TELEWRITE: A NEW TELEHEALTH-BASED ASSESSMENT TO EVALUATE THE HANDWRITING SKILLS OF CHILDREN IN FIRST THROUGH THIRD GRADE

by

Julia M. Guzmán

Dissertation Committee:
Professor Lenin C. Grajo, Sponsor
Professor Lori Quinn

Approved by the Committee on the Degree of Doctor of Education

Date: May 19, 2021

Submitted in partial fulfillment of the
requirements for the Degree of Doctor of Education in
Teachers College, Columbia University

2021

ABSTRACT

TELEWRITE: A NEW TELEHEALTH-BASED ASSESSMENT TO EVALUATE THE HANDWRITING SKILLS OF CHILDREN IN FIRST THROUGH THIRD GRADE

Julia M. Guzmán

Telehealth is needed urgently nationwide, given the COVID-19 pandemic. It is especially urgent in rural and less populated areas where healthcare access is limited. Currently, because there are no pediatric handwriting assessments validated for telehealth use, the TeleWrite assessment would fill an unmet service need and expand the use of telehealth-based occupational therapy (OT) assessment in pediatric practice.

This dissertation explored the preliminary psychometric properties of TeleWrite, a handwriting assessment tool designed to measure the legibility and fluency of handwriting for children in first through third grade administered via telehealth. A series of studies were completed to determine initial interrater reliability, content validity, and clinical utility using classical test theory.

The Rasch model of measurement was used to determine the preliminary psychometric properties of TeleWrite using Winsteps® (v. 4.7.0). The quantitative Rasch analysis of TeleWrite included administration of the tool to 148 children from first to third

grade. This study tested the initial construct validity (internal validity) and test reliability of TeleWrite using the Rasch model of measurement. The Partial Credit Model (PCM) was used for rating scale analysis because TeleWrite is composed of three distinct scales (*handwriting rate, accuracy,* and *fluency*) that differs per task (near point or far point) and per grade level. The Rasch analysis showed a generally good fit with the Rasch unidimensional model, indicating strong construct and internal validity and moderate ability to separate abilities of students reliably in terms of handwriting skills. However, following the Rasch model, a larger sample is necessary to obtain improved calibration, reliability, and validity measures.

This study and supported by the literature described the need for a new handwriting evaluation tool validated for telehealth use. The findings of the current research contribute to the literature and OT practice as the first handwriting assessment specifically designed and validated for telehealth use that assesses all pertinent variables of handwriting associated with handwriting difficulties.

DEDICATION

To my parents who supported me with love and understanding.

To my dearest husband, Noel, my rock and biggest supporter who makes my dreams possible. Thank you for always believing in me and supporting me through all life adventures.

My accomplishments would not mean anything if I did not have Nick and Alan by my side. For all I have achieved in life, nothing makes me prouder than my wonderful, intelligent, and kind children.

ACKNOWLEDGMENTS

I would like to thank Columbia University, Department of Occupational Therapy for affording me the unimaginable opportunity to complete this endeavor despite all the challenges of the past year.

To my graduate students at the University of Scranton who pilot tested the TeleWrite and to all the families who generously allowed their children to be tested for this study.

It is my genuine pleasure to express deep gratitude to my mentor, theorist, and guide, Dr. Lenin C. Grajo for his relentless support. As a teacher and mentor, he has taught me more than I could ever give him credit for here. He has shown me by example what a good researcher should be. Many thanks as well to my committee members, each of whom has provided me patient advice and guidance through the editing process.

I humbly extend my thanks to all who helped shape and reshape this work.

I can because of you.

J. M. G.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

Chapter I

INTRODUCTION

This dissertation is a report of the development and preliminary psychometric properties of TeleWrite, a handwriting assessment tool for children in first through third grade administered through telehealth. A series of studies was completed to determine initial interrater reliability, clinical utility, and content validity using Classical Test Theory. This dissertation presents the use of the Rasch measurement model for rating scales to establish internal validity using goodness of fit and analysis of standardized residuals as well as to establish the tool's test reliability. Additionally, a rating scale analysis was completed to optimize the functioning and sensitivity of the rating scales within TeleWrite.

The first chapter of this dissertation presents the background of the study, specifies the research problem and purpose, explains the theoretical framework for the study, delineates the research questions and hypotheses, describes the significance of the study, offers a brief overview of methodology, and concludes with the presentation of the dissertation.

Background

Telehealth is a service delivery model that increased 960% in rural areas between 2011 and 2018 (Olson, McSwain, Curfman, & Chuo, 2018). Telehealth or virtual care delivery was estimated to reach 7 million health consumers in 2018, with spending

increasing from $15.6 billion in 2014 and predicted to increase to $20 billion by 2020 (Olson et al., 2018). Occupational therapy (OT) services delivered via telehealth have increased significantly since the outbreak of the COVID-19 pandemic and are being recognized and reimbursed by many commercial payers nationally. In response to the pandemic, Congress enacted emergency legislation, the Coronavirus Aid, Relief, and Economic Security (CARES) Act, that enabled the Centers for Medicare & Medicaid Services (CMS) to provide funding and temporarily waive restrictions on the provision of OT and other therapy services via telehealth and allows OT to receive payment for Medicare telehealth services (American Occupational Therapists Association [AOTA], 2020).

Telehealth is needed urgently nationwide, given the COVID-19 crisis. This medium of delivering healthcare services is essential in rural and less populated areas where healthcare access is limited (AOTA, 2020). OT services continue to be necessary to help children with disabilities improve their functional skills and accomplish a variety of school-related daily tasks. With an increased need for reaching rural and underserved populations, telehealth has demonstrated the ability to provide evaluation and services to individuals with minimal access to healthcare without compromising the quality of care (Heimerl & Rasch, 2009; Ruble, McGrew, Toland, Dalrymple, & Jung, 2013; Vismara et al., 2018).

The prevalence of handwriting problems in typically developing children has been estimated to range from 5% to 25% (Zwicker & Harris 2009). The assessment of handwriting using formal, standardized, and valid tools is crucial in providing objective measures and quantitative scores for clinical assessment and monitoring progress in

students' handwriting performance (Feder, Majnemer & Synnes, 2000). An evidence-based review (Collins, 2008) found that most current handwriting assessments are criterion-referenced and scored based on the perceptual judgment of the examiner. In addition, the Evaluation Tool of Children's Handwriting (ETCH) (Amudson,1995) and the Children's Handwriting Evaluation Scale (CHES) (Phelps & Stempel, 1987) lack sensitivity to measure change in handwriting performance and questionable test-retest reliability (Collins, 2008).

The evidence supports the need for new handwriting evaluation tools that can better capture the complexity and multidimensional nature of handwriting skills. Critically as well, better handwriting tools validated for telehealth use are needed. As the first Rasch analyzed handwriting assessment, the TeleWrite has many advantages over traditional pediatric writing assessments. Rasch analysis is the preferred method for instrument development (Hambleton & Jones, 1993). Rasch allows for predictability of participants' responses and the ability to identify unexpected results which has research and clinical applications. Rasch methods can provide evidence that the instrument is measuring its constructs in a way that matches what a theory would predict (McAllister, 2008). In addition, the Rasch model can calculate difference between expected and observed performance scores (McAllister, 2008). As a result, the model may help to overcome shortcomings in currently used handwriting scales regarding lack of discrimination and sensitivity.

The TeleWrite handwriting assessment (Appendix A) is designed for administration at home or school through a web-based platform and follows a consultation model where the occupational therapist observes the child in his or her natural

environment and gathers information from the caregiver and child. The therapist can then explain evaluation results directly to the child and the parent through a virtual consultation and begin to build relationships through collaboration. A web-based assessment format has many advantages over traditional handwriting performance assessments, including convenience, cost reduction, and the delivery of services to remote areas or those with limited therapists (Benham & Gibbs, 2017). The TeleWrite assessment extends evaluative services for children with handwriting difficulty who live in remote areas or with limited access to OT. Assessments were conducted remotely in the children's natural environment of home. This tool is an effective way of providing an occupation-based handwriting evaluation at a significantly reduced cost over standard face-to-face assessment.

Research Problem

Telehealth has become a widely utilized service delivery method across a variety of populations, especially during this challenging time of the COVID-19 pandemic. However, limited information is available for pediatric practice. There is preliminary support for the feasibility, accuracy, and clinical utility of telehealth-based intervention for children (Criss, 2013; Norbakht, Rassafiani, Hosseini, & Ahmadi, 2017), but the reliability and accuracy of telehealth-based pediatric assessments are still underexplored.

The American Psychological Association (APA, 2020) recommends that practitioners refrain from administering assessments via telehealth when these assessments are exclusively validated for face-to-face administration, especially if they require in-person contact, such as handling assessment materials, standardized interactions between examiner and client, and clinical observation of the person in particular contextual factors. Clinicians should not alter the test administration to a

different format and must use caution when interpreting data and making conclusions to inform clinical decisions if the administration protocol is not followed (APA, 2020). Because research and evidence for the parity of testing in a remote or online format compared to a traditional, face-to-face format are limited (Jacobs, Cason, & McCullough, 2015), most OT clinicians in telehealth use observational behavior checklists or interviews for remote assessments. These measures are subjective and not validated to provide objective data or standardized scores of the child being evaluated (Janes, Persch, Schwartz, & Cason, 2016).

Telehealth is important in pediatric practice and is becoming more critical than ever during the COVID-19 crisis. To date, there are no validated, objective, performance-based handwriting assessments for telehealth administration in the literature. Because of the length of the COVID-19 pandemic thus far, there is an urgent need to develop and determine the reliability and validity of OT assessments administered remotely. The TeleWrite purports to meet this need as the first telehealth-based handwriting assessment.

Research Purpose

The purpose of this dissertation was to develop a performance-based handwriting assessment especially designed to be administered through a remote web-based platform and to establish its psychometric properties. TeleWrite was designed to provide early recognition of the symptoms of handwriting challenges in elementary school children. This dissertation addressed four research questions:

Research Question 1: Does TeleWrite define a unified construct of handwriting skills that indicate its internal validity?

Research Question 2: Do the items of TeleWrite (writing rate, accuracy, and fluency) fall within a linear pattern when organized hierarchically, indicating goodness-of fit in the Rasch model of measurement and therefore supporting construct validity?

Research Question 3: Does TeleWrite reliably separate clients in a continuum of increased to decreased handwriting skills and, therefore, indicate test reliability?

Research Question 4: Is the rating scale of TeleWrite optimized so that it sensitively measures varying levels of handwriting skills and, therefore, indicates rating scale responsiveness?

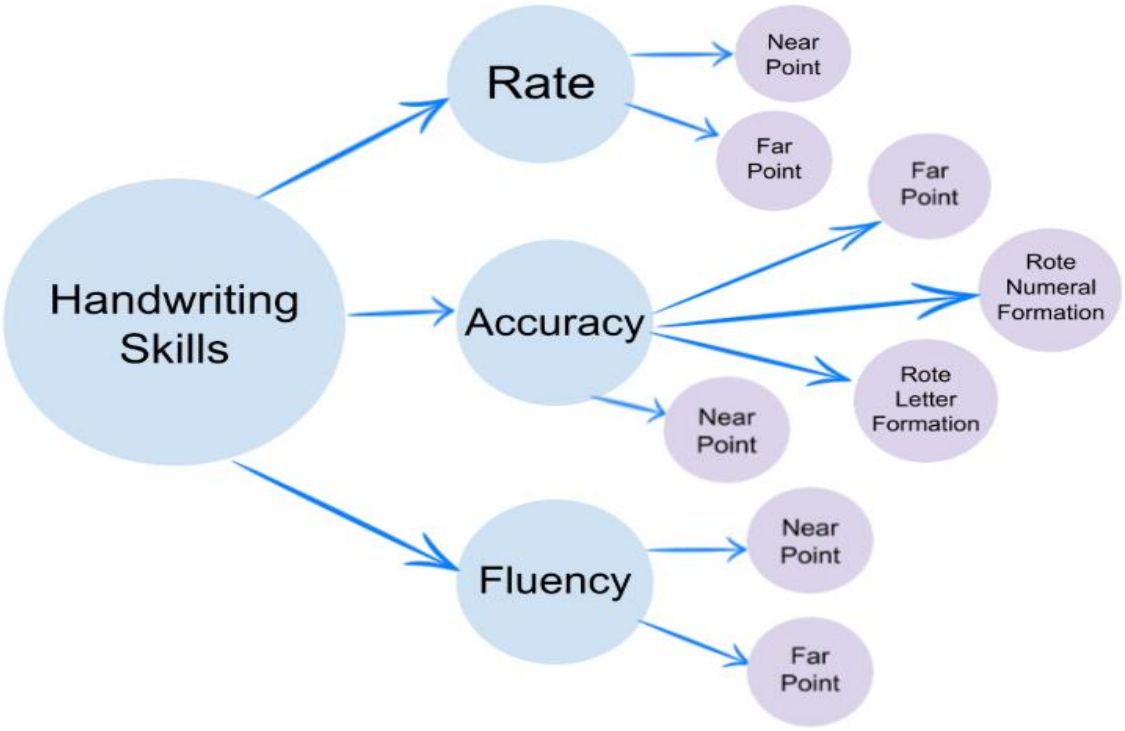
Theoretical Framework

Relationships among Latent Variables of TeleWrite

In assessment development, the underlying construct that a tool is intended to reflect is called the *latent variable* (DeVellis, 2012). A latent variable denotes a characteristic of the construct under examination, which TeleWrite characterizes as handwriting skills. While a latent variable is a true score, it cannot be measured directly (DeVellis, 2012), and certain variables must be observed first to assess the main construct. In TeleWrite, rate, accuracy, and fluency of handwriting are sub-latent variables measured to infer the true value of the latent variable of handwriting skills. Figure1 illustrates the relationship between variables assessed in TeleWrite. It assesses the motor output of letter formation, numeral formation, and handwriting copying skills from near and far point sources (observable variables).

Figure 1

Latent and Sub-Latent Variable Diagram of TeleWrite



TeleWrite was designed as an occupation-based assessment of handwriting skills by examining:

- *handwriting rate* (i.e., the number of letters written by the student per minute);
- *handwriting accuracy* (i.e., as the number of letters written "correctly" per minute);
- *handwriting fluency* (i.e., how fast [rate] a student can correctly [accuracy] copy the letters per minute).

These variables can be observed in the context of far and near distances.

Core Assumptions of the TeleWrite Assessment

There is a lack of handwriting assessments designed specifically to capture information on constructs of handwriting performance from ecological or motor skill models. TeleWrite is a performance-based assessment designed to be administered in the child's natural environment of classroom or home. It assesses both the process and the result of the child's transaction with meaningful contexts while participating in writing tasks. The context comprises all interdependent conditions that surround a child (Dunn, Brown, & McGuigan, 1994). TeleWrite allows therapists to assess the features of context such as the physical, temporal, and social elements of the environment, which are important factors that influence handwriting performance.

Performance Skills

The TeleWrite assessment examines several motor performance skills outlined in the *Occupational Therapy Practice Framework* (4th edition) (AOTA, 2020), including alignment and stabilization of paper, positioning of the child in relation to desk and writing materials, grasp, and in-hand manipulation skills of dominant hand as well as coordination of both hands together to manipulate, hold pencil, and/or stabilize paper while performing writing tasks. In addition, TeleWrite considers performance skills (i.e., motor skills such as stabilizing, positioning, grasping utensils, manipulating, coordinating, endurance, and pacing) and process skills (attends, sequences, organizes). When assessing handwriting, Feder and Majnemer (2007) recommended also assessing necessary client factors such as a stable grip, in-hand manipulation skills, gradation of force, visual perception, motor planning/praxis, visual motor coordination, and visual memory—all components of the TeleWrite assessment. Additional prerequisites for handwriting include

perceptual, motor, cognitive, and language functions, as well as integration of these components (Donica & Holtz, 2019).

Ecological Model

The Ecology of Human Performance (EHP) (Dunn et al., 1994) provides an occupational therapy framework, emphasizing "a person's contexts as a critical variable in his or her ability to perform functional tasks in his or her life" (Dunn, McClain, Brown, & Youngstom, 1998, p. 12). The EHP framework postulates that occupational performance is a product of the interaction among the person (child), the context, the task, and the performance (Dunn et al., 1994). For instance, a negative change in any of the four components disrupts the balance between them and may reduce the person's performance in activities including handwriting (Dunn et al., 1998). The EHP is a framework that considers the relationships among persons (i.e., their skills, abilities, and experiences), what persons want and need to do (i.e., their desired task performance), and where they need to conduct their daily lives (i.e., the contexts for desired performance) (Dunn et al., 1998). Based on this model, changing the person, the context, the task, or the relationship between these influences will cause a change of improvement or deterioration in performance. Function and dysfunction are defined in terms of the clients' effectiveness in achieving their goals (Dunn et al., 1994). Because of the transactive nature of EHP interaction, OT developmental evaluations should include observations of children and their context.

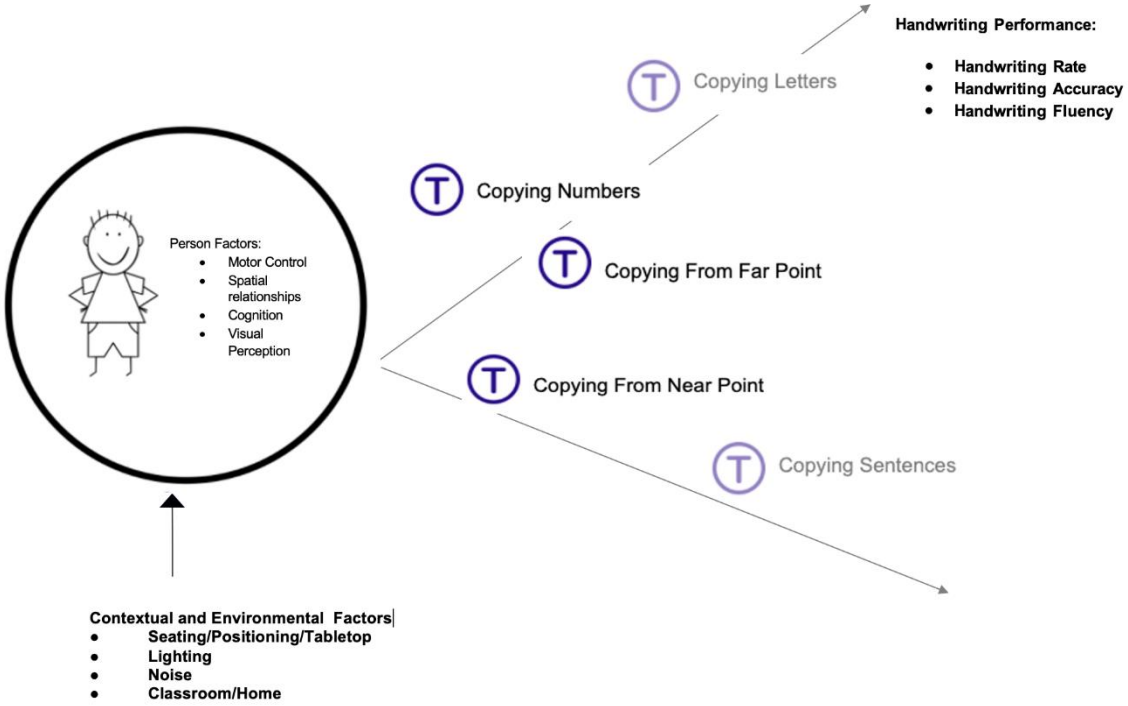
Because the EHP model highlights the context, it is especially suitable to be used for assessment in a natural environment. TeleWrite is designed to be administered in the child's natural environment to provide a holistic picture of the child's handwriting abilities

in the context of school or home. The assessment process typically examines the interaction of the learner with the environment. Assessment in the natural environment therefore includes the observation of specific transactions between parent or teacher and child, the types and quantity of objects available in the environment, specific events that occur in the home or at the direction of the parents, and specific features of the physical environment of the house or classroom (Sangster-Jokić & Whitebread, 2011). Some environmental features observed through the TeleWrite assessment include seating, positioning, noise, and lighting level.

Figure 2 applies principles of the EHP to the theoretical assumptions of the TeleWrite assessment.

Figure 2

Ecological Model Applications to TeleWrite



TeleWrite considers person factors, occupational performance, and contextual and environmental aspects as equally important in determining the handwriting performance range. In evaluation, therapists look at overall handwriting performance through the sub-latent variables of handwriting rate, accuracy, and fluency to assess the underlying deficits that interfere with successful handwriting. TeleWrite describes the interaction between performance components and functional performance in handwriting. For example, during evaluation, therapists will assess several areas of function, including biomechanical and ergonomic factors (e.g., sitting posture, pencil grip); quality of writing (e.g., letter formation, directionality, spacing); and ecological factors (e.g., lighting, noise, and behavioral responses).

According to the EHP model, a person's performance range may be wide (enhanced) or narrow (limited) (i.e., enhanced or limited handwriting performance, rate, accuracy, or fluency) due to the transaction between the person factors (i.e., the child's handwriting skills); the demands of the task (i.e., near point, far point copying); and the contextual and environmental factors (i.e., noise, lighting, etc.). For example, a student's handwriting performance range may be considered enhanced (i.e., wide) when it depicts optimal performance, as noted when a student obtains observed scores that match the student's skills and abilities. The provision of a supportive environment and a handwriting task that offers the "just right challenge" and matches the student's skills will yield a wide range of performance. On the other hand, a student's handwriting performance range may be considered limited (i.e., narrow) and indicative of dysfunction and need for intervention when there is a mismatch between person variables, task demands, and/or contextual features (Dunn et al., 1998). For example, the performance

range may be narrow when a student with limited visual perceptual or fine motor skills is given a handwriting task that is too challenging, even when the context is supportive.

Assessment of children's handwriting traditionally includes evaluation of fine motor, perceptual skills, motor planning, and quality of movement, and effective assessments examine fine motor control, visual motor integration, visual perception, kinesthesia (Feder, Majnemer, & Synnes, 2000; van Hartingsveldt, de Groot, Aarts, & Nijhuis-van der Sanden, 2011) and traditional OT assessments focus on task performance or on person skills, interests, abilities, but few (if any) target the environment or context (Dunn et al., 1998). This is the advantage of developing an ecologically valid and reliable assessment such as the TeleWrite that highlights the interdependent nature between the person and the environment and its effect on handwriting performance.

Handwriting is a complex activity that involves perceptual and motor components. The motor and perceptual components related to poor handwriting performance may include fine motor control, visual motor integration, visual perception, kinesthesia, and sensory modalities (van Hartingsveldt et al., 2011). Writing by hand helps to train the brain to integrate visual and tactile information and fine motor dexterity (Engel, Lillie, Zurawski, & Travers, 2018) and writing letters mediates neural or functional specialization (Francher, Priestley-Hopkins, & Jeffries, 2018). Moreover, the brain develops a functional specialization that integrates sensation, movement control, and cognition to emphasize planning and execution of movement in motor performance (Sangster-Jokić & Whitebread, 2011). Brain imaging studies have revealed that multiple areas of the brain, such as the cerebellum and left dorsal premotor cortex, are core sites for the development of handwriting as they become co-activated during the learning of

writing letters. The right posterior cerebellar hemisphere, the ipsilateral side to the writing hand, is considered a writing-specific motor region (Palmis, Danna, Velay, & Longcamp, 2017).

Motor Learning Theory

Motor learning refers to the acquisition of motor skills or the performance enhancement of learned motor skills or the reacquisition of skills (Magill & Anderson, 2014). Task-specific practice is an approach that focuses on performance of functional tasks that are meaningful to the individual (Muratori, Lamberg, Quinn, & Duff, 2013). Thus, when using this type of practice, a therapist must be able to assess the child accurately and identify the child's limitations and deficits. Children have great variability in motor performance, and this variability could affect learning, retention, and re-learning of letter forms (Musselman, Roemmich, Garrett, & Bastian, 2016).

As described by Gentile (1998), motor skill learning involves two interdependent processes, implicit and explicit, directed toward attaining a functional goal. These processes are reorganized with practice; however, learning occurs at different rates, with explicit learning occurring at a faster rate than implicit learning (Magill & Anderson, 2014). Because explicit learning is directed toward conscious awareness of environmental conditions, therapists are encouraged to use strategies that direct the learner's attention to the regulatory features of the environment, such as demonstrations or verbal instructions. On the other hand, implicit learning requires varying environmental conditions that generate force production and challenge organization during functional activities (Gentile, 1998). Explicit learning "generates verbal knowledge of movement performance, involves cognitive stages within the learning

process and is dependent on working memory involvement" (Kleynen et al., 2015, p. 2). The TeleWrite assessment considers explicit learning during rote learning tasks such as formation of upper- and lowercase letters and numbers, where learners are required to tap into their working memory to complete the rote tasks.

Motor Control

Motor skill acquisition and motor control consist of several stages (Magill & Anderson, 2014). Gentile (1972) suggested there are two objectives for the initial stage of motor skill development: (a) to learn the basic movement pattern needed to achieve the goal; and (b) to identify components of the environment important to the task (Muratori et al., 2013). In the early stage of motor skill acquisition, when using a closed-loop mechanism, the child must engage in active problem solving to find strategies to match features of task and environment. In the closed-loop system, feedback is compared against a standard such as letter forms stored in memory that allow the child to carry out an action or make a correction. The use of a closed-loop feedback control requires several practice sessions to attain automation of movement. In the later stages (i.e., writing fluency), children use open-loop (feedforward) control and automaticity. In fluent handwriting, movement is conducted using open-loop (feedforward) control and attained in an automated manner (Chang & Yu, 2010). The feedforward strategy uses internal representations to pre-plan and anticipate the necessary motor sequence to achieve a motor goal (Rosenbaum, 2009). These movements are subject to a speed/accuracy trade-off because the faster the movement is carried out, the less accurate it becomes. Thus, inefficient writers switch to closed-loop motor control when the demands for accuracy increase (Chang & Yu, 2010).

Handwriting is a hierarchically organized representation of motor movement, the interaction between lower-level perceptual motor (i.e., motor learning, execution) and higher-level cognitive processes (i.e., executive functions). Low-level perceptual motor functions such as spatial positioning of letters, letter formation, and writing letters on a line become automatic, allowing for the activation of higher-level processes (Stievano, Michetti, McClintock, Levi, & Scalisi, 2016).

Motor learning tenets have guided the development of TeleWrite, especially in relation to written work production. TeleWrite describes handwriting competence in terms of handwriting rate (i.e., speed), handwriting accuracy, and handwriting fluency. TeleWrite assesses handwriting autonomy through writing fluency scores. Handwriting fluency is considered the most accurate index of automation (Palmis et al., 2017) as well as the most functional aspect of handwriting ability because the goal of efficient handwriting is to allow children to focus on higher-order aspects of writing or automaticity (Engel et al., 2018). The goal of writing by hand is to promote writing automaticity or to write with minor letter form variation and without conscious awareness (Staats, Oakley, & Marais, 2019). Rate of writing speed is important to assess in elementary school-age children because the rate of writing develops in a rather increasing linear fashion throughout primary school. Handwriting speed is an aspect of handwriting performance that becomes more important as students encounter expectations to produce longer, more complex writing (Stievano et al., 2016).

Overview of the TeleWrite Assessment

The TeleWrite Assessment is an ipsative or self-referenced and descriptive assessment test designed to measure the rate, accuracy, and fluency of children's

handwriting in Grades 1 through 3. Data gathered from the TeleWrite aims to measure the child's present performance, which may be used as a baseline to compare with future development and identify the needs for handwriting skills remediation. The *writing rate* in TeleWrite is calculated based on the number of letters written by the student in minutes; *writing accuracy* is calculated as the number of letters written correctly; and *writing fluency* is calculated based on how fast (rate) a student can correctly (accuracy) copy the letters per minute. TeleWrite assesses handwriting skills in four observable domains: (a) letter formation of lowercase and uppercase letters; (b) numeral writing; (c) near point copying; and (d) far point copying (i.e., distance copying).

Intended Population and Setting

The targeted population included students approximately 6 to 8 years of age and in first through third grade of schooling. Children with or without a diagnosis who have mild or moderate deficits in the areas of visual motor and/or fine motor skills that impact their performance with handwriting skills may be referred by either caregivers and/or teachers for assessment. Grades 1-3 were selected because students reach important motor milestones in the early elementary school years. The quality of handwriting develops quickly during Grade 1 (ages 6-7 years), reaches a plateau by Grade 2 (ages 7-8 years), and sees further development by Grade 3 (ages 8-9 years) in that handwriting becomes automatic, organized, and available as a tool to facilitate the development of ideas (Stievano et al., 2016).

The TeleWrite assessment is designed to be administered in the child's natural environment of home or the classroom, with the assistance of a parent who will receive specific instructions prior to administration (Appendix B). While the TeleWrite

assessment encompasses a systematic, holistic approach that considers the child and the family within the context of their life environments, it is designed to follow Western cultural orthographic rules and to be administered in the English language.

Test Item Selection and Construction

TeleWrite is an ipsative, performance-based tool designed to provide a holistic picture of the child's handwriting abilities in the context of school or home. Ipsative assessments are conducted in the child's natural environment and allow the observation of specific transactions between parent or teacher and child, the types and quantity of objects available in the environment, specific events that occur in the home or at the direction of the parents, and specific features of the physical environment of the house or classroom. The TeleWrite assessment examines rate, accuracy, and fluency of four observable components of handwriting, including alphabet production of lowercase and uppercase letters from memory, numeral writing 0-9 from memory, and near point and far point copying.

The Scoring Scale

The *writing rate* in TeleWrite is calculated based on the number of letters in the writing prompt copied by the student in minutes. To determine the writing rate, the therapist will time the student while copying the letters in the prompt in minutes. The following formulas yield the letters per minute (LPM) writing rate for the child.

Writing Rate = *number of letters in prompt ÷ number of minutes to complete*

Writing accuracy is operationally defined as the total number of letters in the writing prompt that are recognizable within context. The following formula is used to calculate writing accuracy:

Writing Accuracy = *Total number of letters written correctly ÷ Total number of letters in prompt x 100*

Writing fluency in TeleWrite is calculated based on how fast (rate) a student can correctly (accuracy) copy the letters per minute. To calculate the writing fluency, divide the total number of letters written correctly from the "accuracy" score by the "rate" score in minutes using the following formula:

Writing Fluency = *Writing Accuracy Score ÷ Writing Rate Score*

As shown in Table 1, TeleWrite created grade-level appropriate handwriting prompts that use every letter of the alphabet for Grade levels 1-3 for near and far point copying. The writing prompts were designed to not make semantic sense to prevent the examinee from predicting or memorizing the sentence to be copied. The researcher used the Lexile Analyzer (https://www.lexile.com) to develop grade-level appropriate prompts for copying during the assessment. A Lexile Measure is represented by the letter "L" and contains a number on a scale of 200L to 1600L, which indicates an individual's ability to read or represents difficulty in reading a text. The Lexile level indicates the targeted reading level for an individual student (Lennon & Burdick, 2004). There are strong links between the reading and writing systems at the word level (i.e., word recognition, reading comprehension and spelling). Berninger et.al., (2002) found reciprocal relationships between reading and writing at grade levels 1-4.

Table 1

Handwriting Prompts for Near and Far Point Copying Per Grade Level

Grade Level	Near Point Prompts	Far Point Prompts	Lexile Range
1st	"The five boxing wizards jump quickly." (31 letters)	"Max, Jack, and Harvey helped with a big quiz." (34 letters)	410L-600L
2nd	"Jahmal quietly picked winter vegetables and boxes for prizes." (52 letters)	"Jaz and Frederick quietly bought many very wild, exciting props." (54 letters)	610L-800L
3rd	"TeleWrite is a fun way to practice writing by using every letter, A-Z. Some just show once, some I know, and some like X, I question." (105 letters)	"The kindergarten teacher favorably started the school year; he quickly jotted down and explained the classroom rules from A-Z." (105 letters)	610L-800L

Chapter II

REVIEW OF THE LITERATURE

Background Literature on Telehealth

Telehealth or telerehabilitation is a rehabilitation model used by a variety of health disciplines in which therapeutic services are provided to clients in their homes or other environments (Veras, Kairy, Rogante, & Giacomozzi, 2015). According to a 2018 position paper, telehealth is defined by the American Occupational Therapy Association (AOTA) "as the application of evaluative, consultative, preventative, and therapeutic services delivered through telecommunication and information technologies" (p. s69). Telehealth is the official term adopted by AOTA (2018) to describe occupational therapy (OT) services provided through telecommunication technology when therapists and clients are in different physical locations. Alternate terms such as telemedicine, telerehabilitation, and telepractice are used in the literature to describe telehealth. In its position statement, the AOTA asserted that this evolving model of healthcare delivery is promising for every major OT practice area (AOTA, 2018).

This emerging area of practice allows therapists and other health professionals to expand healthcare beyond traditional settings for individuals who live remotely or without direct access to rehabilitation services. While face-to-face interventions still allow for beneficial results for clients, research has shown that providing therapeutic

services to clients remotely can also produce successful outcomes (Benham & Gibbs, 2017; Cason, 2014). Telehealth is a way for healthcare professionals to provide their services to patients in a different physical location which can save time, cut travel expenses, and provide healthcare to areas lacking basic or specialized health services (Cason, Behl, & Ringwalt, 2015).

Telehealth is an expanding field where occupational therapy can have a significant impact, yet several barriers inhibit its implementation as a widely used service delivery model. The literature identified that laws and regulations as well as reimbursement are the leading barriers to the adoption of telehealth (Cason, 2014). One of telehealth's major roles in the U.S. healthcare system is to provide health services to patients living in rural areas lacking healthcare providers. Often, telehealth is conducted across states, requiring practitioners to hold a professional license in the state where the client is located (Cason, 2014). The deliverance of interstate telehealth services is hindered by the current lack of licensure portability laws, which often inhibits practitioners' ability to provide clinical services. The implementation of telehealth across states is complicated since regulations and policies often differ between states (Cason, 2014). AOTA is currently working on creating an interstate professional licensing compact for OT to address licensure portability. The Occupational Therapy Licensure Compact legislation must be passed into law in each state where it will apply. The goal for this multi-year initiative is to begin state participation by 2024 (AOTA 2020).

Another complication in the use of telehealth in the United States is that technologies and software employed for the delivery of OT services via telehealth must meet Health Insurance Portability and Accountability Act (HIPAA) and HITECH (Health

Information Technology for Economic and Clinical Health Act) requirements to assure the client's protected health information is kept private and confidential (Cason, 2014). In March 2020, the U.S. government supported the increase in need for telehealth service delivery with a temporary change in enforcement of the HIPAA of 1996 and the HITECH Act of 2009 regulations associated with security requirements for technology used to deliver healthcare services (U.S. Department of Health and Human Services Office for Civil Rights, 2020).

Reimbursement for telehealth services, especially in OT, has been an issue affecting the expansion and implementation of telehealth. As a result of the COVID-19 pandemic, the Center for Medicare and Medicaid Services (CMS) has temporarily expanded the scope of services approved for telehealth to include OT evaluations. Occupational therapy evaluations of low (30 minutes), moderate (45 minutes), and high complexity (60 minutes) have been temporarily approved for Medicare reimbursement (CMS, March 1, 2020). Despite these barriers, positive trends in telehealth and its implementation into OT practice are emerging. In the face of other professions utilizing telehealth services for many years, this is an emerging area of practice that is increasingly being utilized in the OT profession. It is within the scope of OT to use telehealth for evaluation, intervention, consultation, monitoring, and supervision to offer clients OT-specific services that will help them achieve their meaningful outcomes (AOTA, 2013; Cason, 2014).

Telehealth allows the client to learn skills in the context in which the activity is typically performed, eliminating the need to transfer skills from clinics to the clients' natural environment (Ziegmann, Cole, Lichtenberg, & Brooks, 2001). As the field of

telerehabilitation has progressed and grown, this innovative treatment approach has diversified the healthcare system, allowing professionals in all areas to provide virtual services to their clients. With an increased need for reaching rural and underserved populations, telehealth has demonstrated the ability to provide services to individuals with minimal access to healthcare without compromising the quality of care (Heimerl & Rasch, 2009; Ruble et al., 2013; Vismara et al., 2018).

Telehealth has been used as a means of delivering a variety of health-related information, assessment, and services, including physical, speech, OT, and other disciplines over large and small distances. There has been much discussion on telerehabilitation services, yet little evidence has been found in the literature to describe the applications used for the treatment and evaluation process of children using this service delivery model. Current research is still lacking to develop an effective intervention plan based on a telehealth-based assessment for the pediatric population. A primary concern of traditional healthcare services in the field of pediatrics has been the lack of equal attention to both children and their caregivers, inhibiting optimal outcomes of treatment. Telerehabilitation has offered a progressive approach to this concern by promoting the collaboration between caregivers and their children rather than the traditional practitioner-client relationship. Telehealth allows those providing interventions to take instruction from practitioners and implement them firsthand.

Service Delivery Settings

Telehealth intervention has been delivered to pediatric clients in several environmental contexts, including natural environments, educational settings, and clinical venues. Therapeutic intervention has been provided in a variety of settings, including

early intervention, pediatric private practice, school-based settings, home-based settings, and behavioral health practices. Several articles have mentioned a combination of services provided at home or in clinic or both (Behl et al., 2017; Ferguson, Craig, & Dounavi, 2018; Kuravackel et al., 2018; Little, Dunn, Pope, & Wallisch, 2016; Ruble et al., 2013; Tomlinson, Gore, & McGill, 2018; Vismara et al., 2018; Wainer & Ingersoll, 2015).

Types of Telehealth Intervention and Assessment

Telehealth intervention has been applied to address communication (Ingersoll, Shannon, Berger, Pickard, & Holtz, 2017), psychosocial interventions (Hepburn, Blakely-Smith, Wolff, & Reaven, 2016), behavioral intervention (Knowles, Massar, Raulston & Machalicek, 2017), motor-based (Reifenberg et al., 2017), sensory-based (Gibbs & Toth-Cohen, 2011), and applied behavioral analysis (Heitzman-Powell, Buzhardt, Rusinko, & Miller, 2014). Most intervention studies have focused on children with Autism Spectrum Disorder (ASD), where services delivered included behavioral and diagnostic assessments, educational consulting, self-guidance after video module training, parent-mediated behavioral interventions, and coaching/training of parents or teachers (Ferguson et al., 2018; Heitzman-Powell et al., 2014; Kuravackel et al., 2017; Little et al., 2016; Ruble et al., 2013; Tomlinson et al., 2018; Vismara et al, 2018; Wainer & Ingersoll, 2015).

There is limited research on the utilization of pediatric assessments through telehealth or teleassessment. However, the available evidence has suggested that administration of teleassessments is successful, reliable, and valid. Researchers have found that assessments such as the Functioning Every day with a Wheelchair-Capacity

(Schein et al., 2011) and the Movement Assessment Battery for Children (MABC-2) (Henderson, Sugden, & Barnett, 2007) can be administered through a telehealth platform, indicating that the use of this technology for the purposes of OT can be of great benefit to the profession (Nicola, Waugh, Charles, & Russell, 2018). Hodge et al. (2018) also found that telehealth assessments can be as successful and as reliable as assessments conducted through face-to-face interactions. Moreover, the study found that the use of telehealth could allow for accessibility of assessments and provide services to children who live in remote areas. This allows for an inclusion of services and children will no longer be denied access due to their location. Assessments completed through telehealth also allow for accommodations in parents' busy schedules and without the need for travel; telehealth allows for flexibility in time, making it a more convenient practice compared to face-to-face interactions (Hodge et al., 2018). Two additional assessments used by occupational therapists have been administered in person, while the intervention was delivered through telehealth. Benham and Gibbs (2017) administered the Bruininks-Oseretsky Test of Motor Proficiency, Second Edition (BOT™-2; Bruininks &Bruininks, 2005), while Criss (2013) virtually administered the Print Tool™ (Olsen & Knapton, 2006) as a pre- and post-assessment. Nonetheless, there remains a gap in assessment validation for telehealth practice in occupational therapy.

Handwriting

Handwriting is an important tool for communication. According to Graham (1992), legible handwriting in school-aged children is a key factor in achieving academic success in the education system. Evaluating handwriting is an important area for pediatric therapists to address, as poor handwriting is the most frequent referral for OT services in

the school system (Cermak & Bissell, 2014). According to McHale and Cermak (1992), as much as 30-60% of a child's school day consists of fine motor activities, with handwriting tasks accounting for much of this time. Recently, there has been greater expectation to assess handwriting in the environment and contexts in which it naturally occurs (Robinson & Penman, 2011), where a child naturally participates in handwriting activities in school or at home. Telehealth allows the occupational therapist to observe and assess the child in their natural environment when it may not be otherwise possible.

Table 2 provides a summary of commonly used handwriting assessments in pediatric practice. The most common handwriting assessment tools that are used in the school and clinic setting are the Evaluation Tool of Children's Handwriting (ETCH-M) (Amudson, 1995), the Test of Handwriting Skills-Revised (THS-R) (Milone, 2007), The Minnesota Handwriting Assessment (MHA, Reisman, 1999), and the Print Tool™ (Olsen & Knapton, 2006). Brossard-Racine, Mazer, Julien, and Majnemer (2012) evaluated the psychometric properties of the ETCH-M. This is a criterion-referenced tool used as an in-person handwriting evaluation for Grades 1 to 6 and is offered in manuscript, and cursive. The ETCH-M takes about 30 minutes to complete six writing tasks. This assessment addresses alphabet and numeral writing from memory, near and far point copying, dictation of nonsense words and numbers, as well as short sentence composition. The letters and numerals throughout the assessment are scored based on legibility, the prevalence of displacements or reversals, poor erasure, and performance writing time in seconds. The researchers concluded that the ETCH-M is an effective tool for evaluating handwriting for children of these ages (Brossard-Racine et al., 2012).

The Test of Handwriting Skills-Revised (THS-R) is a norm-referenced tool, commonly used to assess handwriting in children ages 6 to 18 years, 11 months old. The tasks in this test include writing uppercase and lowercase letters from memory in alphabetical order, uppercase and lowercase letters from dictation out of order, single-digit numbers, randomly selected uppercase and lowercase letters, modeled words, and sentences as well as copying a sentence from dictation (Donica & Holt, 2019). This test looks at and scores the number of letters the child can write in 20 seconds, the number of letters reversed, the total letters touching, and the number of lowercase letters written as uppercase letters and vice versa (Milone, 2007).

The Minnesota Handwriting Assessment (MHA) is a norm-referenced standardized assessment that tests a child's handwriting. The assessment takes less than 5 minutes to administer and can be completed in manuscript or print. This test assesses the spacing, size, alignment, and form of the letters and words and overall legibility (Roston, Hinojosa, & Kaplan, 2008). This assessment's psychometric properties validate its effectiveness; the test-retest, intra-rater, and interrater reliability are exceptionally high (Roston et al., 2008).

The Print Tool™ is a non-standardized handwriting assessment to evaluate handwriting of children in Grades 1 to 3 (Olsen & Knapton, 2006). The task involved in this assessment includes printing uppercase letters in alphabetical order, lowercase letters in word format, and single-digit numerals all from dictation. There is little research on the psychometric properties of this assessment, but a study was conducted that showed the test-retest correlation for the Print Tool™ is moderate to good and should be used to guide treatment rather than determine a handwriting deficit (Donica & Holt, 2019).

Table 2

Common Handwriting Measures Used in Pediatric Occupational Therapy

Evaluation Tool of Children's Handwriting (ETCH) Amundson (1995)	A criterion-referenced tool designed to evaluate the manuscript and cursive handwriting skills of children in Grades 1 through 6.
Minnesota Handwriting Assessment (MHA) Reisman (1999)	A norm-referenced, near point copy test. Students are asked to copy words onto a marked lined paper. This assessment helps to identify students with handwriting difficulties through an objective analysis and normative rating system. It can also provide baseline data to document treatment effectiveness by specifically focusing on the students' rate of completing the writing task legibility (speed), alignment, size, and spacing.
The Print Tool™-5th Edition Olsen & Knapton (2006)	A non-standardized printing assessment used in evidence-based remediation programs. The Print Tool™ is used to evaluate handwriting skills, plan intervention, and measure progress in students who are experiencing handwriting difficulty. Eight handwriting components for capitals, lowercase letters, and numbers are considered.
Test of Handwriting Skills-R Milone (2007)	A norm-referenced tool that examines both manuscript and cursive handwriting through dictation, near point copying, and alphabet writing from memory. Normative data are provided for students 6 years to 18 years, 11 months.

Summary

There is ample support for telehealth-based assessment and intervention for children. Some barriers found in the literature included technological barriers and the need to utilize encrypted connections during the telehealth sessions as well as provide training to the professionals involved in the sessions to ensure an effective session with minimal technological disruptions. The research studies included in this review indicated that interventions via telehealth are viable. However, challenges to intervention still exist, such as reaching families who live in geographic isolation, licensure requirements, and extensive time commitments from families. Practitioner advocacy and legislative involvement on behalf of families and children are potential solutions to overcome these barriers. Due to the COVID-19 pandemic, AOTA and state associations are advocating with state governments to ensure that consumers can access OT services.

Telehealth and technology-based applications have the potential to supplement or even replace traditional service models to increase access to evidence-based services, but more research is needed in this area. The evidence supports the use of telehealth across many areas of pediatric rehabilitation. For clients where geographic location, availability of providers, and time commitment are barriers to cost-effective and timely care by rehabilitation professionals, including occupational therapists, telehealth may alleviate these concerns.

This review of the available literature identified several gaps, the most important being the lack of handwriting evaluation tools validated or specifically designed for telehealth administration.

The Development of TeleWrite

The development of a handwriting assessment to implement via telehealth is imperative because of extenuating circumstances (i.e., global health crisis). Four studies were conducted as preliminary studies to develop the TeleWrite assessment.

- Study 1 (Chapter III): A study with eight children ages 6-8 to determine the suitability of the assessment and any need for preliminary test item revision. Preliminary interrater reliability testing used this beta version of TeleWrite with (n = 9) raters.
- Study 2 (Chapter IV): A clinical utility survey of (n = 55) pediatric occupational therapists to obtain their clinical impression regarding the usefulness and accuracy of the beta-version of the TeleWrite assessment.
- Study 3 (Chapter V): A content validity analysis to determine if the test items of TeleWrite have content validity as reported by (n = 9) subject experts. This study informed the revision of TeleWrite.
- Study 4 (Chapter VI): A rating scale analysis completed to optimize the functioning and sensitivity of the rating scales within TeleWrite. Measures of validity were conducted by administering TeleWrite to children (n = 148) in Grades 1-3. This study used Rasch analysis and fit analysis for unidimensionality, rating scale function, to obtain item calibrations or person measures for stability and reliability.

Chapter III

STUDY 1: INTERRATER RELIABILITY

Interrater reliability is the degree of agreement among raters. Interrater reliability has been the primary target of research on test development, and one of the most key features of educational and psychological instruments (Lautamo, Laakso, Aro, Ahonen, & Törmäkangas, 2011). Good interrater reliability guarantees that the instrument can be confidently used across raters. Therefore, evaluating interrater reliability is an essential step in the development and standardizing of an assessment instrument.

Study Design

To measure interrater reliability, the researcher used classical test theory to analyze the intraclass correlation (ICC) estimate of interrater reliability on quantitative data. The ICC is a measure of reliability, specifically the reliability of two or more different raters to measure subjects similarly. Interrater reliability is important as it validates that a scale is robust to fluctuations in raters. Hence, scales with high interrater reliability are less prone to measurement error caused by variation in human judgment (Bobak, Barr, & O'Malley, 2018).

Participants

The participants were a convenience sample of eight children between the ages of 6 and 8 years. This age group was selected since children are developing the foundations

of handwriting at this time. The participants included three boys and five girls; three were 6 years old, two were 7 years old, and three were 8 years old. All children were right-handed and completed the entire assessment via telehealth. In addition, eight graduate assistants were recruited to assist the principal investigator with the administration and interrater scoring of the assessment.

Inclusion criteria. The following inclusion criteria were used to determine participant selection: typically developing children or children reported by their families or teachers as having mild to moderate learning disabilities included in a general education classroom; first, second, or third graders approximately between the ages 6 to 8 years old and English-speaking.

Exclusion criteria. The following exclusion criteria were used to determine participant selection: children with severe developmental delays as determined by their ineligibility to be included in a general education classroom, and children not familiar with the English alphabet and/or not English-speaking.

Data Collection

After receiving full committee approval from the Internal Review Board (IRB) of the University of Scranton, the TeleWrite assessment was administered via a secure Zoom® virtual platform to each of the eight participants by eight graduate students and the principal investigator (PI). The training of the graduate assistants by the PI followed a process established by Castorr et al. (1990). The training consisted of 4 hours each week for a period of 4 weeks in three phases:

1. training graduate student raters to use the instrument by administering the TeleWrite assessment to one another through a web-based platform;

2. evaluating rater performance at the end of training; this was done by re-scoring an earlier version of the TeleWrite assessment and comparing new and old scores; and
3. determining the extent to which rater training was maintained during a reliability study (Castorr et al., 1990). This was confirmed during statistical analysis that yielded high levels of interrater reliability coefficients.

Participants received the TeleWrite packet and consent forms in the mail and the assessment was completed through a secure Zoom® platform. The nine raters (eight graduate research assistants and the PI) administrated the tool and rated the eight children in eight different sessions, each lasting approximately 1 hour over a period of 4 weeks. Behavioral observations were made during the assessment and the handwriting prompts were scored upon receipt of packets in the mail. All raw data were recorded over a period of 3 weeks. There was no communication between each of the raters during the scoring process to ensure the veracity and integrity of the scoring process.

Data Analysis

The principal investigator coded all nine ratings in a spreadsheet which was reviewed for accuracy by one of the graduate students. The Statistical Package for the IBM (SPSS) Statistics for Windows®-Version 25.0 was utilized to calculate the interrater reliability among the nine raters. Internal consistency with Cronbach's alpha (α) statistics evaluated the interrelatedness of items to the construct of interest (Bobak et al., 2018). Intraclass correlation coefficients (ICC) were calculated for spatial relations and rote letter formation.

Results

The interrater reliability for total scores yielded a score of 0.928 (excellent) at a 95% confidence interval, and the Cronbach's α measure for internal consistency was 0.953 (excellent). The intraclass correlation coefficient for sub-domain scores was between good and excellent. Table 3 provides a summary of ICC of the TeleWrite sub-domains.

Table 3

Intraclass Correlation Coefficient (n = 9)

Total Measures	0.928
Spatial Relations	0.882
Capital Letters	0.917
Numerals	0.912
Lower Case Letters	0.857
Cronbach's α	0.953

Chapter IV

STUDY 2: CLINICAL UTILITY

An assessment is considered to have adequate clinical utility when it is deemed clinically feasible and useful in terms of length of administration, cost, and ease of administration (Schwartz, Averbuch, Katz, & Sagiv, 2016) based on perception by potential test users. Clinical utility is usually determined by surveying clinicians who have reviewed and possibly administered the assessment in their clinical setting (Fawcett, 2009). This study examined the clinical utility of the TeleWrite assessment.

Study Design

This study used a mixed-method approach by collecting quantitative descriptive demographic data and qualitative data through the survey results of occupational therapists' perceptions of the clinical usefulness of the instrument.

Participants

The participants were recruited through a convenience sample utilizing social media platforms, including CommunOT (AOTA's social media platform) and various OT Facebook groups such as Pediatric Occupational Therapists, School-based OT, Teletherapy for SLPs and OTs, and Research4OT.

Inclusion and Exclusion Criteria

The following inclusion criteria were used to determine participant selection: English-speaking occupational therapists who reside in the United States or overseas, currently licensed and practicing pediatric occupational therapy, at least 3 years of experience administering pediatric handwriting assessments. Participants who did not met the inclusion criteria were excluded from participating in this study.

Data Collection

Clinical utility data were collected from occupational therapists who viewed the social media posts and responded with interest in participating in the clinical utility study. Participants were instructed to reach out via email and were emailed instructions on how to complete a Qualtrics® survey. Participants were then emailed a link to a 10-minute instructional video that explained how to administer the TeleWrite Assessment through a secure platform that complies with HIPAA requirements. After indicating consent, the participants were asked to review a beta version of the TeleWrite assessment, optionally test it out to students/clients, and complete a survey about their perceptions of the usefulness of the tool in practice. The survey consisted of 17 demographic, clinical usefulness, and open-ended questions (Appendix D).

Data Analysis

The data collected from the clinical utility survey were analyzed for descriptive information such as frequencies and percentages in the survey software Qualtrics®, and later exported to SPSS® v25 for statistical descriptive analyses. Initially, 128 pediatric occupational therapists responded as being interested in the survey, and 55 occupational therapists completed the survey, with a 69% response rate.

Qualitative data were analyzed using a collective case study methodology. This methodology analyzes a common set of questions designed to understand the individual experiences of the participants and gain experiential insights (Balog, 2016) by combining common themes and experiences as collective cases. This methodology was used to understand the participants' individual and collective experiences when examining the TeleWrite assessment. The participants were asked to respond to Likert-type questions and comment with depth and contextual detail.

To establish trustworthiness in qualitative data analysis, three qualitative confirmability methods were used following Lincoln and Guba's (1986) criteria for evaluating qualitative research. First, an audit trail was instituted to establish that the findings were based on participants' responses instead of the researcher's own preconceptions and biases. The data from the open-ended responses in the survey were independently gathered and highlighted for significance by six graduate students who assisted in the construction and distribution of the survey. Second, cross-comparisons on the analysis were made. The primary researcher and mentor formed individual clusters of themes from the highlighted statements of the participants' responses to each question. The themes were then coded and cross-analyzed for accuracy by the two researchers. Lastly, a participant checking protocol was implemented. This ensured that the responses and comments the participants provided accurately represented their collective sentiments about the TeleWrite assessment. The themes were emailed to all 55 participants to verify the accuracy of the themes. There were no objections and participants agreed to the veracity of the findings.

Results

Demographic Information

The participants of this study represented 12 different countries (see Table 4). The United States accounted for the highest number of participants from 22 different states. Most participants (60%; n = 33) in this survey had more than 16 years of experience as an occupational therapist. The primary practice area of most of the participants was in schools (78%; n = 39). The majority (71%; n = 79) of participants declared no telehealth experience at all.

Table 4

Clinical Utility Demographic Data

Country of Residence (n = 56)	n
Australia	1
Bangladesh	1
Belgium	1
Croatia	1
India	1
Israel	1
Lebanon	1
Netherlands	1
New Zealand	1
Pakistan	1
United Arab Emirates	1
United States of America	45
50 States, DC, and Puerto Rico (n = 42)	**n**
Arkansas	1
California	3
Colorado	3
Delaware	2
Florida	1
Idaho	1
Illinois	3
Kansas	1
Massachusetts	2
Michigan	3

Table 4 (continued)

50 States, DC, and Puerto Rico (n = 42)	n
Minnesota	1
Missouri	1
New Hampshire	2
New Jersey	1
New York	7
North Dakota	1
Ohio	3
Oklahoma	1
Pennsylvania	1
Texas	1
Washington	2
Wyoming	1
Years of Experience (n = 55)	**n (%)**
3-5 years	5 (5.8)
6-10 years	1 (6.8)
11-15 years	11 (12.5)
More than 16 years	33 (37.5)
Primary Practice Setting (n = 55)	**n (%)**
Academia	2 (2.6)
Early Intervention	4 (7.2)
Outpatient	1 (1.8)
Other	5 (9.0)
Schools	43 (78.1)
Telehealth Experience (n = 55)	**n (%)**
No Experience	39 (70.9)
1 or More Years	16 (29.1)

After analyzing the data in IBM SPSS® version 25, the responses to the following questions were obtained (Table 5). The data from the responses showed that (61%,

n = 35) of the participants indicated they were quite or extremely likely to use TeleWrite in their practice. Similarly, there was a favorable response regarding the perceived accuracy of TeleWrite in assessing handwriting skills. Most of the participants (64%; n = 34) indicated that TeleWrite was quite/extremely accurate in measuring handwriting performance.

Table 5

Clinical Utility Descriptive Data

"Ease of Instructions" (n = 55*)	n (%)
Not Easy	1 (1.82)
Moderately Easy	13 (23.64)
Quite/Extremely Easy	41 (74.54)
"Accuracy of Measuring Handwriting Performance" (n = 53*)	**n (%)**
Not Accurate	0
Moderately Accurate	19 (35.8)
Quite/Extremely Accurate	34 (64.15)
"Likelihood to Use the TeleWrite Assessment in Practice" (n = 52*)	**n (%)**
Not Applicable	2 (3.70)
Not Likely	3 (5.56)
Slightly Likely	6 (11.11)
Moderately Likely	10 (18.52)
Quite/Extremely Likely	33 (61.11)

*n = varies per question

Qualitative themes. There were many positive clinician perceptions on the clinical utility of TeleWrite. Participants provided valuable information and useful feedback. The results of this survey provided insight into the ease, usefulness, and accuracy of this assessment tool in pediatric OT practice. The respondents indicated that the TeleWrite assessment has the potential to be a useful handwriting screening tool for pediatric occupational therapists who need to administer assessments through telehealth. The qualitative data demonstrated a strong positive response on the clinical utility of the tool.

Theme 1. TeleWrite is easy to use and has an easy-to-follow format. Most of the respondents described TeleWrite as easy to follow. Several respondents commented on the user-friendly format for administration and scoring.

> "I like the format, easy to follow and assesses all that I would want from a handwriting assessment whilst being age appropriate."

> "It looks really comprehensive, and I like the easy scoring."

> "I quite liked that this assessment combines the observations I have been doing for years into a neat form!"

> "I think it is a great assessment. It can be completed fairly quickly and is easy to follow and score."

> "We need a good handwriting assessment in Telehealth, so I think this is great!"

Theme 2. TeleWrite is a potentially useful tool in practice. A shared constructive response from most participants indicated that they might use TeleWrite in their own practices.

> "I would love to try this in my practice."

> "Overall looks like a nice tool to have as a school-based OT."

> "As a school-based OT in schools for 11 years...we are trying to move away from standardized tests. This is a great way to segment a functional look of the mechanics of writing. I think either in the classroom or virtually this would be a helpful assessment."

Theme 3. Need for clearer examples and clearer instruction. Some of the participants suggested that TeleWrite may be strengthened and more useful for clinical practice if it had illustrations for pencil grasp and a script for instructions and materials needed for administration.

> "An instruction sheet to clearly define terms/explanatory pictures for sections such as pencil rotation and grasp would benefit therapists who might not be as familiar with these observations."

> "You might want to include pictures of different pencil grasps."

> "A script for exact instructions. Is there a template that the family should be given, and a list of utensils to gather before the session?"

Theme 4. Potential to include other observable variables. Other observable variables were suggested by the participants to enhance the tool.

> "I think there should be a dictation piece of single words to mimic a spelling test and a composition section where they compose 1-3 sentences based upon their grade."

> "From what I have seen, the assessment assesses transcription and does not include independent writing."

> "Maybe look at finger opposition and use of alternate hand as an assist."

> "In this form there is no sensory issue part of hands."

This preliminary research project provided valuable information on the clinical utility of the TeleWrite assessment. The results of this survey provided insight into the ease, usefulness, and accuracy of this assessment tool in pediatric OT practice. According to the participants, the TeleWrite assessment has the potential to be a useful handwriting screening tool for pediatric occupational therapists utilizing telehealth.

Chapter V

STUDY 3: CONTENT VALIDITY

Content validity refers to the process of ensuring that the items of the assessment tool represent all facets of a given construct (Hand, Darragh, & Persch, 2018). According to the COnsensus-based Standards for the Selection of Health Status Measurement INstruments (COSMIN Standards) (Mokkink, Prinsen, Bouter, Vet, & Terwee, 2016) there are four requirements for appropriate content validity:

1. All items should refer to relevant aspects of the construct to be measured.
2. All items should be relevant to the study population (e.g., age, gender, disease characteristics, and setting).
3. All items should be relevant to the measurement instrument (discriminative, evaluative, and/or predictive).
4. All together, items should comprehensively reflect the construct to be measured (Ailliet, 2013).

The most widely used method of quantification of content validity is the content validity index (CVI), which is derived from the rating of the content relevance of the items of an instrument (Lynn, 1998).

Study Design

The purpose of content validity is to determine the representativeness and relevance of the items of an instrument to the construct of interest (Zamanzadeh et al.,

2015). Content validation serves as an important indicator of the instrument's quality (Polit, Beck, & Owen, 2007) and as a link between abstract theoretical constructs and measurable indicators (Wynd, Schmidt, & Schaefer, 2003). The item-level content validity index (I-CVI) was calculated to measure the strength of each item from the experts' ratings. The calculation of scale-level content validity index (S-CVI) is the proportion of content-valid items. The I-CVI yields item-level information that can be used to refine or discard items of the scale, while the S-CVI is a summary of the content validity of the overall scale (Polit et al., 2007). Both I-CVI and S-CVI produce the proportion of agreement of the expert panel on the relevance of the items of a scale. This is done by averaging the I-CVIs to compute total the scale-level or S-CVI (Polit et al., 2007).

Participants

After obtaining approval from the Institutional Review Board (IRB) of Teachers College, Columbia University, IRB Protocol Number 20-223, this study recruited a panel of participants deemed to be content area experts. This project assessed the content validity of TeleWrite through a panel of nine experts. The criteria for an expert were an individual who has at least 10 years of professional practice in pediatrics, at least one professional publication, multiple professional presentations, or experience with or participated in creating and/or validating assessments. This study required content expertise; thus, exclusion criteria included all those not meeting the inclusion criteria above. As per Lynn (1998), a minimum of five reviewers will minimize chance agreement and the total number should not exceed 10. The content expert list was generated from a hand-search of authors of frequently read or cited publications on

handwriting assessment or intervention from the *American Journal of Occupational Therapy* and the *Journal of Occupational Therapy, Schools, and Early Intervention*. A short list of 10 experts was created. The participants were then recruited via email invitation and asked for consent to participate. All 10 experts consented to participate, and nine experts completed the content review.

Data Collection

Using an online survey platform, the expert reviewers were sent a copy of the TeleWrite assessment and asked to review the assessment materials. The expert reviewers then utilized a four-point ordinal scale to assess relevance for each item as (1) not relevant, (2) minimally relevant, (3) moderately relevant, and (4) highly relevant following guidelines by Zamanzadeh et al. (2015). A section for comments on each test item was provided so experts can provide suggestions on how to further strengthen the test item.

Data Analysis

This study analyzed two content validity indices—item level content validity index (I-CVI) and scale level validity (S-CVI)—which were calculated by dividing the number of experts who scored an item a 3 (moderately relevant) or 4 (highly relevant) by the total number of experts. The inventory was dichotomized by combining values 1 and 2 together (not relevant) and values 3 and 4 together (relevant) for each item. A content validity index (CVI) of 0.80 indicates good content validity, and if the S-CVI is greater than or equal to 0.90, the scale has excellent content validity (Zamanzadeh et al., 2015).

The data were computed by entering the formulas on Microsoft Excel for Windows (version 16.26, 2019). Content validity was determined by calculating for item and scale content validity indexes (I-CVI/S-CVI) from experts' ratings. The researcher used the following formulas to calculate the I-CVI and S-CVI (Polit et al., 2007).

$$\textbf{I-CVI} = \frac{(n \text{ experts that scored 3}) + (n \text{ experts that scored 4})}{\text{total } n \text{ of content experts}}$$

$$\textbf{S-CVI} = \frac{(\text{sum of all I-CVI scores})}{\text{total } n \text{ of scored items}}$$

In addition, the researcher calculated the modified kappa (*k*) statistic to analyze the probability of chance agreement among content experts utilizing the method recommended by Wynd et al. (2003). This method to compute a modified kappa statistic

$$k = \frac{P_o - P_c}{1 - P_c}$$

(*k**) can be used to verify I-CVI for agreement among multiple experts that the item is relevant. The probability of chance agreement increases with the number of raters from diverse backgrounds (Wynd et al., 2003). This study had a high number of expert reviewers (n = 9) and utilized the modified kappa (*k*) statistic to reduce the probability of agreement beyond chance agreement. Kappa values range from +1.00 to –1.00, with a positive kappa indicating interrater agreement occurring more frequently than would be expected by chance. A +1.00 demonstrates complete agreement across raters (Wynd et al., 2003). A zero kappa indicates that agreements are no more than can be expected by chance. A multi-rater modified kappa coefficient of 0.74 and above is considered excellent; 0.60 to 0.73 is considered good; and 0.40 to 0.59 is considered fair (Wynd et al., 2003).

Results

Nine content experts with an average of 20.7 years of experience in pediatric OT or academia (range 8-40+ years) completed the review. All content experts had obtained a doctoral degree (research doctorate or PhD n = 7; post-professional OTD or non-research doctorate n = 2). Most of the experts (n = 7) indicated having 16+ publications related to pediatric OT practice and all had more than 16 workshops, posters, short courses, or in-service presentations related to handwriting, telehealth, or pediatrics. The content experts indicated that all 10 TeleWrite subscales rated (in-hand manipulation, near/far point copying, rote letter/rote numeral copying, near point writing prompt, far point writing prompt, spatial relationships, timing formula, writing rate formula, accuracy formula, and fluency formula) were considered essential items, as indicated by an I-CVI of 0.80 or higher. The total scale content validity index S-CVI was 0.94. Table 6 contains a summary of all ratings. The (*k*) value for all raters varied from 0.88 to 1.0, which indicates excellent concurrence among the experts beyond chance agreement and further validates the I-CVA and S-CVI findings.

Table 6

Content Validity Index for the TeleWrite Assessment Tool

TeleWrite Test Item	Highly Relevant	Moderately Relevant	Minimally Relevant	Not Relevant	I-CVI	Modified *k**
In-Hand Manipulation	3/9	5/9	1/9	0/9	1.00	1.00
Near Far Copying	4/9	4/9	1/9	0/9	0.88	0.88
Rote Letter Rote Numeral	9/9	0/9	0/9	0/9	1.00	1.00
Near Point Writing Prompt	7/9	2/9	0/9	0/9	1.00	1.00
Far Point Writing Prompt	7/9	1/9	1/9	0/9	0.88	0.88
Spatial Relationships	8/9	1/9	0/9	0/9	1.00	1.00
Timing Formula	4/9	4/9	1/9	0/9	0.88	0.88
Writing Rate Formula	6/9	2/9	1/9	0/9	0.88	0.88
Accuracy Formula	8/9	1/9	0/9	0/9	1.00	1.00
Fluency Formula	7/9	1/9	0/9	0/9	0.88	0.88
					S-CVI	0.94

Four Themes

The open-ended questions provided qualitative feedback from content experts that can be summarized into four themes.

Theme 1. The variables in TeleWrite are all important aspects of handwriting to assess. The content experts rated the sub-latent and observable variables of the assessment as salient aspects of handwriting to measure. Examples of feedback include:

> "This section is well done. I like the simplicity in scoring and the percentage rating at the end to give a concrete score on how the child did with upper case, lower case, and numbers."

> "This is good. The instructional note regarding example for first grade prompt was helpful. Perhaps a completed example for each grade level in the instruction guide would be helpful."

Theme 2. The formulas for scoring appropriately measure writing rate, writing accuracy, and writing fluency. According to feedback from the experts, the measurement scales and formulas used are suitable to calculate the constructs under examination. Examples of feedback include:

> "I like that the minutes are determined first. It makes it easier to determine the writing rate by just plugging in the minutes."

> "I like the scoring structure of this section and the score guide (2, 1, 0)."

> "This is good. A reference for grade level equivalency of legible letters per minute for both near and far point copying would be beneficial to include in the instructional guide for easy reference."

Theme 3. Need for more objective definitions for behavioral observations. Some content reviewers offered valuable observations to help improve the clarity of the tool.

> "These are good tasks for pencil skills, maybe add something about fatigue at the end of the copying."

> "You might consider providing illustrations for the pencil grasps so that you can improve interrater reliability."

> "Will the child be copying from a standard phrase? Depending on the age of the child, the number of fixations could change, and it could be due to their reading level."

Theme 4. Need for grade-appropriate writing prompts and writing paper.

Several expert raters asked for a revision of sentences for copying and writing materials.

> "I would simplify the language of these prompts. Children may have difficulty understanding the prompt, which may affect their writing legibility/rate."

> "The use of four-lined paper could be confusing for the near point prompts. I am not sure how often 3rd grade students are still using that type of paper."

> "I'm not sure why you are using guidelines for 3rd grade, or even 2nd grade, writing. Often it is the ability of the child to manage the spatial demands of writing that is a problem. I don't think they use that type of paper in 3rd grade."

The content review revealed that all subscales of TeleWrite are essential aspects of handwriting. However, the researcher decided to revise to enhance the instrument. Some of the revisions included the modification of the item "number of visual fixations" to a descriptive observation because the raters indicated that it is difficult to quantify the number of visual glances though a video screen. In addition, based on the feedback from the experts, the writing prompts were revised for appropriateness and simplicity, and included punctuation marks for copying. The type of writing paper was also modified to be more specific and appropriate for each grade level (1st, 2nd, or 3rd grade). Overall, the content validation process with the opportunity for reviewers to comment provided helpful insight to modify or clarify some of the item criteria and to make decisions about retaining or revising items from the TeleWrite assessment. The revisions were the basis for the validation version of TeleWrite.

Chapter VI

STUDY 4: CONSTRUCT VALIDITY, RELIABILITY, AND RATING SCALE FUNCTIONING

This study describes the process of determining the internal validity and test reliability of the TeleWrite handwriting assessment.

Research Design

Validation measures that have been shown to provide evidence of fairness, reliability, and item fit measures were used to support the validity of the TeleWrite assessment for its intended use. This study proposed to analyze the initial construct validity (internal validity) and test reliability of TeleWrite using the Rasch model of measurement. There are two commonly used measurement theories in the field of psychometrics: classical test theory (CTT) and Rasch analysis. CTT was utilized to test for content validity and interrater reliability. Rasch analysis, originally considered a special case of item response theory, was used to determine internal validity.

CTT is a commonly used method for evaluating the psychometric properties of an instrument. CTT is compatible with traditional statistical programs, based on correlation and regression methods, that yield universally recognized psychometric properties (e.g., validity, reliability, internal consistency, etc.), terms that are easily understood by most researchers and consumers of research. CTT has three fundamental scores: the observed or test score, the true score, and the error score (Hambleton & Jones, 1993). CTT has been developed to quantify measurement error and to solve problems such as correcting

observed dependencies between variables due to measurement errors. The true score is generally defined as the expected value of a person's observed score if the person were tested an infinite number of times on an infinite number of equivalent tests. Therefore, the benefit of the true score is that it reflects the stable characteristic of the object of measurement (i.e., the person) (Fawcett, 2009). The primary assumption of CTT is that a test score (observed score) is the sum of a person's true score and the measurement error, and these can be interpreted against a reference or norm group (Allen & Yen, 2002). CTT assumes a normal distribution of the sample collected; all items are of equal difficulty and there is no matching between item difficulty and respondent's abilities. Similarly, CTT does not account for scalability and assumes that the size of the jump between rating categories is equal (Allen & Yen, 2002). While widely accepted, CTT has some limitations in terms of instrument validation. First, the test data are expected to follow a normal distribution to be useful with parametric tests, and non-parametric tests require additional analyses (Linacre, 1994). However, with smaller sample sizes typically used when validating assessments using CTT, the assumption of normal distribution is not necessarily met. Second, CTT measures are entirely sample-dependent, which means that the psychometric analysis is based on the total scores of the sample, which may result in internal consistency issues depending on the ability or variability of performance of the sample of examinees (Magno, 2009). Unless studies are conducted with very large samples with representativeness (e.g., gender, age, and other social variables), caution must be made when interpreting validity with small sample sizes. Lastly, tests that produce ordinal scales such as surveys or checklists can only be analyzed using descriptive statistics such as mean, percentages, or effect size. CTT requires interval level

(equidistant scaling) for level of significance or correlation testing (*p* or *r* values) (Magno, 2009).

The Rasch model of measurement is traditionally clustered under the umbrella of item-response theory. Rasch provides three reliability measures for complete assessment tools: person, person separation, and item reliability (McAllister, 2008). Unlike CTT, when the data fit the Rasch model, they follow the principle of invariance to indicate that the test measures the same construct in the same way across subgroups of respondents. When the data fit the Rasch model and the principle of invariance is met, the tool does not require a strict parallel distribution of the sample to determine test validity and reliability (Smith & Suh, 2003). Rasch measures are independent of which items and which persons participate in the measurement process because of this principle of invariance.

In contrast to CTT, Rasch also focuses on item scores rather than total sample test scores. Rasch uses an item reliability index ranging from 0-1.0 with better reliability closer to 1.0, which is analogous to CTT's Cronbach's alpha scale. Lastly, unlike CTT that assumes all items of a test have equal difficulty and the distance between interval ratings is the same, Rasch can use raw scores or ordinal-level measures and convert them to equal interval-level measures that can be easily analyzed using parametric measurements (Boone & Noltenmeyer, 2017).

A limitation of Rasch is that it requires an equal number of persons and test items. For example, a 30-item test requires at least 30 examinees to make the test calibration stable, typically resulting in a larger sample requirement than CTT (McAllister, 2008). Rasch makes similar assumptions with CTT about the observed and true scores, but it

delves more deeply into modelling the relationship between responses to each item and each construct (Baghaei, 2008), allowing for more advanced statistical analyses. While CTT allows for group comparisons in terms of significance or effect size, Rasch can explain the meaning of those differences (Boone & Noltenmeyer, 2017).

Rasch has been used to analyze performance in a linear scale that accounts for level of difficulty of test items. That is, an item exhibiting higher difficulty than the ability level of the respondent will have a lower probability of being correctly answered than an item of difficulty below the ability level of the respondent (Boone, 2016). This property is called Targeting. Rasch has been used extensively in instrument development to measure clinically significant outcomes (Boone & Noltenmeyer, 2017). Test constructs range on a continuum from lower to higher levels of construct, and if the data deviate greatly from the continuum, they are considered misfitting and should either be removed or revised (Magno, 2009). Estimates of item difficulty have been shown to be stable across samples and test forms, providing superior internal consistency and construct validity (Grajo, Candler, Bowyer, Schultz, & Thomson 2018).

Research Questions and Hypotheses

The following research questions were addressed in this study.

Internal Validity

Research Question 1: Does TeleWrite define a unified construct of handwriting skills that indicate its internal validity?

One measure of construct validity is internal validity. Rasch uses the process of the test of unidimensionality to assess construct validity by assessing whether all test

items support a single latent variable. The advantage of measuring one construct at a time is the ability to control (calibrate) for underlying factors that may confound the results (Smith, 2002). Underlying factors that may threaten the construct validity of a tool are referred to as construct-irrelevant variance (Bond & Fox, 2015). These are sub-dimensions that are irrelevant to the main construct (i.e., handwriting) but produce reliable (reproducible) variance in test scores, thus reducing test stability (Smith, 2002).

To determine the internal validity of TeleWrite, the researcher ran an analysis of standardized residuals and principal components analysis. To satisfy the unidimensionality assumption in the Rasch measurement, the unexplained variance in the first contrast/first secondary dimension should have an eigenvalue of less than 2.00 (Linacre, 1998). A principal components analysis was subsequently applied to the residuals to determine whether the domains constituted a unidimensional structure (i.e., whether there were any additional dimensions present). An eigenvalue < 2 for the first contrast (i.e., once the variance explained by the Rasch structure has been factored out) and > 50% of the variance explained by the Rasch structure are indicative of a unidimensional structure (Smith, 2002). Thus, this study set the standardized residual variance explained by measures to at least 40% and the eigenvalues on the first loading to be ≤ 2.0 logits (units of measurement).

Item Targeting and Other Supports for Construct Validity

Research Question 2: Do the items of TeleWrite (writing rate, accuracy, and fluency) fall within a linear pattern when organized hierarchically, indicating goodness of fit in the Rasch model of measurement, and therefore supporting construct validity?

Test items range on a continuum from lower to higher levels of the construct. Fit statistics represent the difference between expected and observed performance scores as predicted by the model and indicate an acceptable degree of variation (Linacre, 2010). To determine goodness of fit, Rasch produces three fit statistics: the first is *item difficulty* or estimates of the chance that the examinees will endorse the concept represented by each item in the scale. The second statistic is *standardized Z statistics (*ZSTD) and *mean square(MNSQ)* values. ZSTD shows the significance of the data: if the data did fit the model, the expected value is 0.0; less than 0.0 indicates too much predictability and more than 0.0 indicates lack of predictability (Linacre, 2010). MNSQ values indicate the amount of distortion of the measurement system with an expected value of 1.0. MNSQ values near 1.0 indicate little distortion of the measurement system, and values greater than 1.0 indicate unpredictability such as noise or data underfit (Linacre, 2010). The third fit statistic, *person reliability*, is a concept similar to Cronbach's alpha or a measure of internal consistency (Linacre, 2010).

The present study used three fit statistics, including Infit and Outfit Mean-squares (MNSQ), person reliability, and Z Standard scores (ZSTD) to report how predictable or how well the levels on each continuum fit the construct of handwriting skills. Outfit MNSQ values are sensitive to unanticipated observations on items that are relatively easy or very hard, such as guessing or simple mistakes (Bond & Fox, 2015), and Infit MNSQ values are sensitive to the pattern of reactions to items directed to the person and vice versa, such as Guttman patterns (Linacre, 2002). Standardized MNSQ values greater than 2.0 logits may indicate distortion in the measurement system (Grajo et al., 2019). Following Linacre's (1994) guidelines, an acceptable range of fit statistics is 0.80 to 1.2,

with high or low fit statistics representing abnormalities in the response pattern to the item. The researcher set Infit and Outfit (MNSQ) values for assessments scored from clinical observations to the 0.5-1.7 logit range and ZSTD score of <2.0 logits (Linacre, 1994) to support goodness of fit of test items.

Test Reliability

Research Question 3: Does TeleWrite reliably separate clients in a continuum of increased to decreased handwriting skills, and therefore indicate test reliability?

To determine test reliability of TeleWrite, the researcher ran an analysis of the separation reliability of the tool. The replicability of a person's ordering and item placement along the continuum on the test determines reliability (Bond & Fox, 2015). An analysis of separation reliability of TeleWrite was conducted to determine the relative ability of the tool to separate persons (Wright, 1996a) based on their levels of handwriting ability (i.e., the latent variable). For person reliability, a reliability value of 0.8 indicates strong ability to distinguish high and low performers using the measurement (Wright, 1996a).

Rating Scale Analysis

Research Question 4: Is the rating scale of TeleWrite optimized so that it sensitively measures varying levels of handwriting skills and therefore indicates rating scale responsiveness?

To determine the functioning of the rating scales of TeleWrite, the researcher ran a rating scale analysis using Rasch-Andrich Threshold step calibrations (Andrich & Marais, 2019). The analysis of step calibrations allows for optimization of rating scale-type tests. This study followed the set guidelines by Linacre (2002) to optimize rating

scale observations: (a) for each rating scale category, there must be at least 10 observations for more accurate log-ratio estimation; (b) category distribution must be as regular as possible to signal lack of aberrant category usage; (c) average measures must advance monotonically with each category; (d) outfit MNSQ values must be less than 2.0 logits; and (e) step calibrations must advance by at least 0.5 logits for 10-item rating scale tests, similar to the guidelines followed by Grajo et al. (2019).

Participants and Recruitment

The participants were recruited through three mechanisms: (a) flyers posted on social media sites; (b) invitations to occupational therapists, also via social media, who served as study liaisons; and (c) invitations via flyers to families within the PI's network.

- Mechanism 1 and 2: Social Media – The PI posted a flyer on the social media website of the American Occupational Therapy Association (https://communot.aota.org/home) and OT Pediatric Facebook groups (https://www.facebook.com/groups/thetelehealthot/; https://www.facebook.com/groups/80351866792/). These groups are accessed by families who are recipients of OT services and occupational therapists.
- Mechanism 3: Personal Network – The PI has worked extensively with families and schools over many years. The researcher reached out to families and personal networks via email and shared a flyer about the study. Once interest to participate was established through the various recruitment mechanisms, the PI emailed the consent and assent form for the parent and/or child to complete. Once the consent and assent forms were received, the PI

scheduled a date and time with the family to conduct the handwriting assessment via telehealth.

Participants were a convenience sample of 148 children between the ages of 6 and 8 years. Parents gave consent to participate and children between the ages of 7 and 8 provided assent to participate. According to Linacre (1994), useful item calibration is the approximation of what can be expected to maintain a useful level of measurement stability. This means that with 100 participants or more, the researcher can be confident that the results will be useful in determining the internal validity of TeleWrite. The Rasch requirements to establish stable item calibration are symmetric; there should be as many items as person measures for a stable item calibration and 10 observations per category (Linacre, 1994). TeleWrite has rating scales with 10 categories; therefore, 100-person measures or more (10 per category) were needed at a minimum to provide statistically stable measures. To obtain item calibrations or person measures stability within ± ½ logit at the 99% confidence level, the sample size should range between 108-243 participants. This sample size meets the minimum criteria in the Rasch analysis to obtain 99% confidence at ±1/2 logit item calibration stability (Linacre, 1994).

Inclusion criteria. The following inclusion criteria were used to determine participant selection: typically developing children or children self-reported by their families or teachers as potentially having handwriting difficulties; English-speaking first, second, or third graders between the ages 6 to 8 years whose parents are willing to assist during the administration of the assessment; and had a computer with a video camera and audio capabilities with stable internet access.

Exclusion criteria. The following exclusion criteria were used to determine participant selection: children with severe developmental delays as determined by their ineligibility to be included in a general education classroom; and children not familiar with the English alphabet and/or not English-speaking or without access to video chatting technology or limited internet connection.

Data Collection

The principal investigator was assisted by 10 graduate student assistants who received a minimum of 10 hours training in administering TeleWrite. The entire assessment including instruction and administration took 30 minutes or less and scoring took an additional 30 minutes. After signed parental consent and verbal assent from participants aged 7 and over was received, every family received a packet, sent via standard mail, containing the TeleWrite assessment and written instructions to access the web-based platform and a self-addressed stamped envelope to return the test forms to the researcher. The test administration was delivered via a HIPAA-secured version of the Zoom platform, licensed through Columbia University Irving Medical Center. The date and time of test administration were individually scheduled by the researcher and the participants. Data was collected during the midst of the COVID-19 pandemic when students were receiving school instruction mostly from home. The participants were asked to choose a typical space in their home where they do school-based tasks to complete the test. The study received Institutional Review Board approvals from both Teachers College and the University of Scranton.

The TeleWrite assessment required the use of testing and scoring forms, a primary or standard pencil, a grade-appropriate writing paper provided in the packet, any writing adaptations that the child commonly uses (e.g., pencil grips), and a comfortable and appropriate-sized desk and chair. The electronic equipment consisted of a laptop or desktop computer with internet access (preferably broadband high internet access) and a web camera either built-in or mounted on the desktop computer. The test packet was sent either through standard mail or electronically to the participants to be received at least 2 days prior to the scheduled assessment. During the scheduled assessment session, the researcher reiterated the instructions to the parent prior to the assessment and the child completed the assessment while live-streamed.

Tool Scales and Administration

TeleWrite is a performance-based handwriting assessment consisting of ordinal-level observations. The observation scales examine functional observations related to use of visual aids (glasses), hand dominance, pencil grasp, stabilization on paper, sitting posture, and analysis of in-hand manipulation skills. The performance-based scales examine rote letter formation of uppercase letters, lowercase letters, and numerals. The observation of spatial relationships in sentence composition assesses writing rate, writing accuracy, and writing fluency. Total scores are calculated for each of the scales (i.e., rote letter formation, sentence copying, and spatial relationships). All items are added and divided by the maximum number of possible points to obtain percentage ratings for each subcomponent. Table 7 presents a summary of the primary and secondary domains of TeleWrite.

Table 7

TeleWrite Primary Performance-Based and Secondary Descriptive Domains

Primary Performance Based Domains		
Domain	Tasks	Scoring System
Rate	Student is asked to copy a grade-level sentence. Therapist records the time to complete task in minutes.	Formula yields a raw score which is converted to a 10-point scale indicating speed of writing.
Accuracy	Therapist scores the writing sample.	Formula yields an accuracy percentage score which is converted to a 10-point accuracy scale.
Fluency	Therapist calculates fluency scores based on writing rate and accuracy.	Formula yields a fluency score which is converted to a 10-point fluency scale.

Secondary Descriptive Domains		
Domain	Tasks	Scoring System
In-hand Manipulation	Child is asked to pick up a coin using the fingertips and bring from palm to finger and finger to palm. Child is asked to rotate a pencil. Child is asked to "walk" fingers up and down the shaft of the pencil.	Observations are recorded using criteria: Skill is Present, Skill is Emerging (partial skill is present), or Skill is Absent.
Spatial Relations	Therapist reviews the student's near point and far point written samples.	Observations are recorded using criteria: Score 0: Skill is observed in less than approximately 25% of the sample. Score 1: Skill is observed in 25%-75% of the sample. Score 2: Skill is observed in more than 75% of the sample.
Rote Letter and Numeral Formation	Child is asked to copy uppercase, lowercase letters, and umbers 0-9 from model.	Criteria: Score 0: Letter is not recognizable out of context or is reversed. Score 1: Letter is somewhat recognizable in context. Score 2: Letter clearly resembles the model.

To administer TeleWrite, the researcher directed the parent with instructions and observed the environment with minimal obtrusion through a webcam, including the setup of the computer and desk in relation to the child, the tools and the materials involved, and the social or physical context in which the writing assessment occurred. Prior to test administration, the family or school was mailed a packet containing two envelopes. One envelope included the writing prompts, grade-appropriate writing paper, written instructions for the parent assisting during testing, and a coin for in-hand manipulation assessment. This first envelope was labelled "Do Not Open until Instructed." The second envelope was a self-addressed stamped envelope to be used by the parents to mail the writing samples back to the examiner for scoring. Parents were also given the option of sending the forms as scanned documents via encrypted email.

During the assessment, the child and parent were instructed to position the computer and video camera at least 2 feet away from the child; the child was seated in a comfortable position where the therapist could observe the child's full body movements during writing tasks; the parent was positioned beside the child during the assessment.

The researcher made observations of the child's ability to participate and succeed in the writing activity and observations of the context of how it supported or hindered the child's performance. The researcher guided the child through the three phases of the assessment (spatial relationships, letter formation, and sentence copying) while the parent provided stand-by assistance. Instructions for assessment activities and materials were then imparted to the parent facilitator. This is the collaborative phase of the assessment in which the working relationship between the parent and the researcher is enhanced by

working on tasks carefully and collaboratively. Record forms and writing samples, once mailed back to the researcher from parents, were kept in a locked file and locked in the researcher's office. Participants were assigned a number code linked to their initials, and data were only reported in aggregate to ensure participants' confidentiality. All messages sent through the platform including text, images, or documents were available only to the researcher conducting the study.

Rasch Data Analysis

The researcher ran several analyses using the Rasch Partial Credit Model (PCM, Masters, 1982) to determine internal validity, test reliability, and rating scale functioning as guided by the research questions. After the raw scores from the TeleWrite administration were entered in a spreadsheet codebook, a test specification file was created and entered on Winsteps® (Version 4.7.0), Beaverton, OR, for Rasch analysis.

The Rasch model is well suited for instrument validation because it follows a hypothetical line where test items are measured against the examinee's ability level called test for goodness of fit (Baghaei, 2008). The Rasch model for rating scales follows a "goodness of fit" model where test items are aligned along this unidimensional, hypothetical line according to difficulty from least to most difficult, and those test items that fall closest to this line are considered to contribute to the construct being observed (InFitting) (Baghaei, 2008). With Rasch, it is possible to evaluate whether test items all measure the same constructs (unidimensionality), whether the items exhibit interval-level measurement (scalability), whether any of the items exhibit bias related to gender (differential item functioning), and whether the items are arranged from least difficult to most difficult (hierarchal ordering of scale items) (Brown & Unsworth, 2009).

Researchers can use raw or scale scores to analyze performance in a linear scale that accounts for unequal difficulties across all test items. The Rasch model allows for corrections by converting raw or scale scores to equal interval measures (Boone & Noltenmeyer, 2017).

The Rasch model uses a vertical ruler that allows for predictions regarding examinees' likely responses; that is, those test items that fall below the ability level of the examinee are more likely to be answered correctly, whereas those test items that fall above the examinee's ability level are more likely to be incorrect (McAllister, 2008). Person's ability and item difficulty are measured in units called logits (log-odd units) ranging from -3 to +3, with 1.5 logits considered acceptable, 2.0 logits good, and 3.0 logits excellent (Boone, 2016).

The Partial Credit Model (PCM) was used for rating scale analysis because TeleWrite is composed of three distinct scales (*handwriting rate, handwriting accuracy, and handwriting fluency*) that differ per task (near point or far point) and per grade level. The PCM allows partial credit to be assigned to specific responses to specific items. Thus, instead of assigning incorrect (no credit) or correct (full credit) to the item final response, the item can be broken down into components. The PCM allows the components of the response to be described from "not at all correct" to "completely correct"; consequently, each item in the instrument can have different scale categories (Smith & Smith, 2004). PCM is widely used for analyzing polytomous items or items for which the responses are scored according to three or more categories (Smith & Smith, 2004). Models of greater complexity and more parameters such as PCM require larger

sample sizes to achieve a reduction in the number of outliers, root mean squared deviation (RMSD), and Type I error reduction (Custer, 2015). TeleWrite is a complex, polytomous instrument consisting of three distinct scales per task and grade level of 10 categories each. For test items to be calibrated with more stringency individually in an instrument like TeleWrite, a sample size of 400 observations to produce equating errors below 0.04 logits is recommended (Custer, 2015).

Results

Participants included 148 English-speaking children between the ages of 6 to 8 years. There were slightly more males, 54 % (n=80) males than females (n=68; 46%). The TeleWrite assessment was administered to 48 first graders, 50 second graders, and 50 third graders with 95% (n=142) recruited from the United States and approximately 5% recruited internationally. Majority had typical handwriting development (81%; n = 120) compared to 19 % (n=28) with self-reported handwriting difficulty. Consistent with the literature approximately 90 percent of the sample was right-handed, whereas the remaining 10% are left-handed or mixed handed (Wühr & Ansorge, 2019). Table 8 presents a summary of the demographic data of the participants.

Table 8

Demographic Data of Participants

Sex	n (%)
Males	80 (54.0%)
Females	68 (45.95%)
Grade Level	**n (%)**
First Grade	48 (32.43%)
Second Grade	50 (33.78%)
Third Grade	50 (33.78%)
Handwriting Difficulty	**n (%)**
Handwriting Difficulty	28 (18.92%)
No Handwriting Difficulty	120 (81.08%)
US/Non-US	**n (%)**
US	142 (95.95%)
Non- US	6 (4.54%)
State or Country of Origin	**n (%)**
California 3 (2.03%)	Connecticut 10 (6.7%)
Delaware 1 (0.67%)	Florida 1 (0.67%)
Georgia 2 (1.35%)	Maryland 3 (2.03%)
Missouri 1 (0.67%)	New Hampshire 2 (1.3%)
New Jersey 54 (36.4%)	New York 29 (19.6%)
Pennsylvania 35 (23%)	Virginia 1 (0.67%)
Philippines 4 (2.70%)	Kenya 2 (1.35%)
Handedness	**n (%)**
Right-handed	121 (81.76%)
Left-handed	25 (16.89%)
Mixed handedness	2 (1.35%)

This study conducted data analysis using the Rasch Partial Credit Model (Linacre, 2002). Useful item calibration is the approximation of what can be expected to maintain a useful level of measurement stability. This means that with approximately 150 participants, the researcher can be confident that the results will be useful in determining the internal validity of TeleWrite. To obtain item calibrations or person measures stability within ± ½ logit at the 99% confidence level, the sample size should range between 108-243 participants, with a suitable sample size of approximately 150 participants (Linacre, 2002). However, to calibrate the data using the Partial Credit Model more successfully, particularly in the rating scale analysis, data simulations recommend a sample size of 400-500 participants (Custer, 2015). The Rasch model of measurement following an instrumentation quantitative study design was used, and the researcher investigated the preliminary psychometric properties of TeleWrite using Winsteps® (Version 4.7.0), Beaverton, OR, for Rasch analysis.

Hypothesis Testing

Research question 1. *Does TeleWrite define a unified construct of handwriting skills and therefore indicate its internal validity?*

Hypothesis. We hypothesized that the observed standardized residual variance would be at least 40% and the raw variance explained by item loading eigenvalues of the first contrast will be ≤ than 2.0 logits, which would indicate lack of divergence issues and, therefore, support internal validity of the constructs.

Internal Validity

The researcher used a principal components analysis (PCA) of Rasch residuals to test for internal validity of TeleWrite using Winsteps® (4.7.0).

The researcher generated an analysis of standardized residuals to explain the variances in the data. The data also generated an analysis of unexpected responses by the students to the items of TeleWrite that may indicate distortions in the data. The dimensionality PCA analysis simultaneously analyzed the interaction between multiple variables that may potentially impact the construct validity of the tool. The residual value (expressed as standardized residuals) is the difference between Rasch model's theoretical expectation of item performance and performance calculated for that item in the data matrix (Bond & Fox, 2015). The criteria for the PCA of response residuals state that at least 40% of the variance should be explained by the Rasch model, and the first contrast in the residuals explains <5% of the variance in eigenvalue units (Linacre, 2002). The PCA of the residuals provides information on possible multidimensionality in the data. For dimensionality analysis using Rasch analysis, we were concerned about the variance explained by the first contrast in the residuals. If this value is large, then there is a second dimension at work (Smith, 2002).

The first hypothesis of this study stated that TeleWrite would be considered unidimensional and internally valid when the variance of the standardized residuals was at least 40% and the variance explained by item loading values of the first contrast was ≤ 2.0 logits after analysis of residuals. The data were calibrated for each grade level for all three domains of TeleWrite: *Rate, Accuracy, and Fluency* in one analysis. The results indicated that eigenvalues >2.0 were found in the observed unexplained variance of the TeleWrite scales for 1st, 2nd, and 3rd grade for all three domains in a single multidimensional analysis (Table 9). To investigate this unexplained variance, the researcher examined the data of the analysis of residual values and found that the

accuracy scores were potentially affecting data convergence. The researcher cleaned the data file by removing all *accuracy* responses to find the impact of these responses on variance in data and the standardized residuals of the observations and ran a second analysis. When the accuracy data were removed, there was an increased variance in the data explained by Rasch, which was a positive finding. This second iteration (Table 9) demonstrated that the *accuracy* scales were causing excessive "noise." The observed values for all scales of 1st, 2nd, and 3rd grades improved to >40% when the accuracy measures were removed.

A principal component analysis PCA of Rasch residuals by Wright (1996b) was used to investigate whether more than one variance component /was explaining the structure of respondent data. Wright postulated that if the data are unidimensional, then components in the residuals will be at the noise level. The test for unidimensionality showed a clear increasing trend from the unidimensional data and the unexplained variance in the first contrast exceeded the 2.0 logit threshold, demonstrating that TeleWrite is potentially comprised of three distinct dimensions, one dimension for each of the three test domains of *rate, accuracy, and fluency*.

To prove this hypothesis, the researcher ran a third analysis: separate calibrations for each grade level separating each of the three domains of TeleWrite. The results of this analysis are shown in Tables 10, 11, and 12. When three separate calibrations for Grades 1-3 for each of the three TeleWrite domains were conducted, the percent of data variance improved significantly (between 50-90% of variance explained) and the first contrast loading values were all under 2.0 logits. The separate calibration results indicated fit with the unidimensionality model of Rasch and showed no convergence issues. Therefore, our

hypotheses were refuted: TeleWrite is a multidimensional tool, consisting of three separate domains. Each of the three domains, as seen in the separate calibration analysis, showed strong internal validity.

Table 9

Multidimensionality Standardized Residuals and Principal Component Analysis

1st Grade (All Three Domains)	Eigenvalue	Observed
Raw variance explained by measures	5.4767	47.7%
Raw variance explained by persons	2.8106	24.5%
Raw variance explained by items	2.6661	23.2%
Unexplained variance in 1st contrast	2.2101	19.0%
1st Grade Multidimensionality with Accuracy Removed	**Eigenvalue**	**Observed**
Raw variance explained by measures	3.1329	43.9%
Raw variance explained by persons	2.8665	40.2%
Raw variance explained by items	0.2664	3.7%
Unexplained variance in 1st contrast	1.9935	27.9%
2nd Grade (All Three Domains)	**Eigenvalue**	**Observed**
Raw variance explained by measures	2.1131	26.0%
Raw variance explained by persons	0.3044	3.8%
Raw variance explained by items	1.8087	22.3%
Unexplained variance in 1st contrast	2.4286	29.9%
2nd Grade Multidimensionality with Accuracy Removed	**Eigenvalue**	**Observed**
Raw variance explained by measures	3.4385	46.2%
Raw variance explained by persons	0.8237	11.1%
Raw variance explained by items	2.6147	35.2%
Unexplained variance in 1st contrast	1.8406	24.7%

Table 9 (continued)

3rd Grade (All Three Domains)	Eigenvalue	Observed
Raw variance explained by measures	6.8684	53.4%
Raw variance explained by persons	4.1492	32.2%
Raw variance explained by items	2.7192	21.1%
Unexplained variance in 1st contrast	2.5037	19.5%
3rd Grade Accuracy Removed	**Eigenvalue**	**Observed**
Raw variance explained by measures	6.7297	62.7%
Raw variance explained by persons	3.8305	35.7%
Raw variance explained by items	2.8992	27.0%
Unexplained variance in 1st contrast	2.0129	18.8%

Table 10

Separate Calibrations Unidimensionality Test for Rate (Grades 1-3)

1st Grade Rate	Eigenvalue	Observed
Raw variance explained by measures	8.4801	80.9%
Raw variance explained by persons	8.1544	77.8%
Raw variance explained by items	0.3256	3.1%
Unexplained variance in 1st contrast	0.0259	0.2%
2nd Grade Rate	**Eigenvalue**	**Observed**
Raw variance explained by measures	5.2310	72.3%
Raw variance explained by persons	4.3793	60.6%
Raw variance explained by items	0.8517	11.8%
Unexplained variance in 1st contrast	0.0278	0.4%
3rd Grade Rate	**Eigenvalue**	**Observed**
Raw variance explained by measures	10.4225	83.9%
Raw variance explained by persons	9.1222	73.4%
Raw variance explained by items	1.3003	10.5%
Unexplained variance in 1st contrast	0.0879	0.7%

Table 11

Separate Calibrations Unidimensionality Test for Accuracy (Grades 1-3)

1st Grade Accuracy	Eigenvalue	Observed
Raw variance explained by measures	6.9325	77.6%
Raw variance explained by persons	5.2116	58.3%
Raw variance explained by items	1.7209	19.3%
Unexplained variance in 1st contrast	0.4174	4.7%
2nd Grade Accuracy	Eigenvalue	Observed
Raw variance explained by measures	22.9688	92.0%
Raw variance explained by persons	8.7544	35.1%
Raw variance explained by items	14.2144	56.9%
Unexplained variance in 1st contrast	0.0545	0.2%
3rd Grade Accuracy	Eigenvalue	Observed
Raw variance explained by measures	4.7952	70.6%
Raw variance explained by persons	4.1710	46.7%
Raw variance explained by items	1.6242	23.9%
Unexplained variance in 1st contrast	0.4244	6.2%

Table 12

Separate Calibrations Unidimensionality Test for Fluency (Grades 1-3)

1st Grade Fluency	Eigenvalue	Observed
Raw variance explained by measures	2.1395	51.7%
Raw variance explained by persons	0.6296	15.2%
Raw variance explained by items	1.5099	36.5%
Unexplained variance in 1st contrast	0.0068	0.2%
2nd Grade Fluency	Eigenvalue	Observed
Raw variance explained by measures	6.8704	77.5%
Raw variance explained by persons	5.8077	65.5%
Raw variance explained by items	1.0627	12.0%
Unexplained variance in 1st contrast	0.0001	0.0%
3rd Grade Fluency	Eigenvalue	Observed
Raw variance explained by measures	4.0013	66.7%
Raw variance explained by persons	3.8059	63.4%
Raw variance explained by items	0.1953	3.3%
Unexplained variance in 1st contrast	0.0031	0.1%

Research question 2. *Do the items of TeleWrite (writing rate, accuracy, and fluency) fall within a linear pattern when organized hierarchically, indicating goodness-of-fit in the Rasch model of measurement, and therefore supporting construct validity?*

Hypothesis. When organized hierarchically, we hypothesized that *"handwriting rate"* is the variable that more easily endorses our latent variable of handwriting skill, while *"handwriting fluency"* is the variable that more stringently endorses handwriting skill. We also hypothesize that all test items of the primary domains of TeleWrite will show goodness-of-fit with the Rasch model.

The second hypothesis states that the items of TeleWrite would fall within acceptable logit measures of Infit and Outfit, mean square (MNSQ) values that indicate goodness-of-fit of the test items with the Rasch model of measurement. To answer this research question, the researcher conducted a goodness-of-fit analysis using Winsteps®. Fit statistics is a quantitative indicator of the accuracy of a measurement tool and, therefore, construct validity. Fit statistics are commonly reported in two forms: mean square values (MNSQ) and standardized Z scores (ZSTD). MNSQ is the mean of the squared residuals for an item, and ZSTD values report the statistical significance of the MNSQ statistics or equivalent to the *t*-test in Rasch analysis (Bond & Fox, 2015). A standard error (SE) quantifies the precision of a measure or an estimate. Unlike in CTT where an SE is given for the sum or total score for all items in a test, in Rasch, an SE is generated for each test item improving test precision and, therefore, test validity. In general, ZSTD scores greater than 2.0 indicate great distortion in the measurement system (Bond & Fox, 2015).

This study used Infit and Outfit mean square fit statistics, SE measures, and ZSTD to examine how well individual items fit the Rasch model. Following Linacre's (1994) values for assessments scored from clinical observations, acceptable Infit and Outfit (MNSQ) values were set to the 0.5-1.7 logit range and ZSTD score of <2.0 logits were used to support unidimensionality of test items.

Table 13 provides a summary of the results of the fit analysis for the three subscales of TeleWrite: rate, accuracy, and fluency from near point (NP) and far point (FP) distances for Grades 1-3. The results support the goodness-of-fit hypothesis for all constructs (NP Rate, FP Rate, NP Accuracy, FP Accuracy, NP Fluency and FP Fluency). All Infit/Outfit MNSQ values fell within the clinically acceptable range of 0.5-1.7 logits. Further, all MNSQ and ZSTD scores fell within Linacre's (1994) more stringent range for Infit and Outfit (MNSQ) values of 0.6-1.4 logit range for rating scale type tests. As shown in Table 13, ZSTD scores of <2.0 logits were also supported, demonstrating further support of evidence of the unidimensionality of the test items of TeleWrite.

Person and Item Hierarchy and Targeting

Additionally, the Rasch analysis provides a vertical ruler to visualize the ability level of the samples in the study juxtaposed to the item difficulty of the TeleWrite test items. The Rasch person-item map is used to compare the range and position of the person measure distribution on the left-hand side to the item measure on the right-hand side. Persons represented in the map as an "x" appear in ascending order from the bottom of the figure to the top. Items on the right are represented by item numbers. A vertical ruler indicates that the closer the items are to 0 logit value, the better fit in the Rasch model (Linacre, 2010). Figures 3, 4, and 5 illustrate the placement of student handwriting

abilities in the NP, FP Rate; NP, FP Accuracy; and NP, FP Fluency categories in the Rasch model of measurement expressed in logits. The vertical ruler also indicates the ability of the participants and the capacity of the TeleWrite items to assess different handwriting ability levels.

Person ability of the sample. The vertical ruler shows samples with lower and higher writing abilities indicating a good range of abilities and most person abilities close to the item difficulties of TeleWrite.

Item difficulty. The more the person's ability surpasses the item's difficulty level, the greater the positive (+) difference and the higher is the person's probability of success. But when the item is too difficult for the person, then the difference is negative (-) and the person's probability of success is less than 0.5 logits. The more difficult the item is for the person, the greater this negative difference becomes and the lower is the person's probability of success (Wright, 1996a). A "logit" scale is used to express item difficulty on a scale that extends from negative infinity to positive infinity. For many analyses, item difficulties will range from -3 logits to +3 logits (Wright, 1996a). The item difficulty levels of TeleWrite all hover in the -1 +1 logit values, indicating an appropriate level of item difficulty for the person ability of the sample.

Targeting. The vertical ruler depicts both person measures and item measures on the same linear scale to allow an analysis of how well the test items are distributed with regard to the ability level of the persons (Boone, 2016). Researchers can evaluate how close the mean item measure is from the mean person measure. When the mean items and mean persons are very close to each other, the assessment has good test-item targeting (Boone, 2016). Given the spread of person abilities and the item difficulty, it seems to

indicate that TeleWrite is best able to assess average abilities rather than extremely low and extremely high abilities most reliably. This may affect separation reliability, which addresses Research Question 3.

Research question 3. *Does TeleWrite reliably separate clients in a continuum of increased to decreased handwriting skills, and therefore indicate test reliability?*

Hypothesis. The reliability coefficient of TeleWrite will be 0.80 or better indicating good ability to separate clients in a continuum of increased to decreased handwriting skills.

To answer the third research question on the reliability of the TeleWrite tool and its ability to separate clients in a continuum of increased to decreased handwriting skills, an analysis of the person and item separation reliability and separation index were analyzed (Bond & Fox, 2015). According to Linacre (2010), a person separation index of at least 2 and reliability of 0.80 is often the benchmark for the practical utility of an assessment tool. Low person separation (< 2; person reliability < 0.80) with a relevant person sample implies that the instrument may not be sensitive enough to distinguish between high and low performers and, therefore, more items may be needed.

This study conducted a Winsteps® analysis of person separation reliability of the TeleWrite scales (Table 13) to determine its relative ability to separate persons based on their levels of handwriting ability, our latent variable. For the data to fit the model adequately, it is generally recommended that the two fit statistics range from 0.72 to 1.30 logits (Bond & Fox, 2015).

The Winsteps® 4.7.0 User Manual (Linacre, 2020b) states that the analysis yields two separate indices, as shown in Table 13. RMSR is the root-mean-square residual that summarizes the differences between observations in category and expectations and Model RMSE, which is computed on the basis that the data fit the model, and that all misfit in the data is merely a reflection of the nature of the model. Real RMSE is computed on the basis that misfit in the data is due to departures in the data from model specifications and reports a lower limit to the reliability of measures based on this set of items for this sample (Linacre, 2020). Separation indices of 3 or greater are desirable; a person's separation reliability index of 1.50 is acceptable; 2.00 is good and 3.00 is excellent (Wright, 1996a). A separation statistic of 0.80 or higher indicates that with the relevant sample, the instrument has adequate test items to sensitively distinguish high and low performers based on the latent variable (Boone & Noltenmeyer, 2017) and therefore support test sensitivity.

The results of the Rasch separation analysis indicated that the person separation indices for the *rate* and *fluency* dimensions for Grades 1-3 were near/within the acceptable range. However, the *accuracy* scales for Grades 1-3 showed low separation indices indicating that TeleWrite can only reliably assess a narrow range of a participant's performance. Similarly, in the present analysis, all person separation reliability values were at or near acceptable levels, except for the *accuracy* reliability scores for scales which were 0.47, 0.60, and 0.20.

Table 13

Fit Analysis and Person Separation Reliability of Grades 1-3

Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
1st Grade Near Point Rate	-.27	.30	.81	-.70	.86	-.44
1st Grade Far Point Rate	.27	.33	.94	-.15	.98	.03
Real: RMSE: 1.67	S.D.: 2.97		Separation Index: 1.77		Reliability: .76	
Model: RMSE: 1.38	S.D.: 3.11		Separation Index: 2.25		Reliability: .80	
Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
2nd Grade Near Point Rate	.21	.25	.85	-.53	.87	-.39
2nd Grade Far Point Rate	-.21	.27	1.00	.10	1.03	.20
Real: RMSE: 1.23	S.D.: 1.78		Separation Index: 1.45		Reliability: .68	
Model: RMSE: 1.08	S.D.: 1.72		Separation Index: 1.75		Reliability: .75	
Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
3rd Grade Near Point Rate	.46	.46	.86	-.34	1.26	.63
3rd Grade Far Point Rate	-.46	.43	.76	-.75	.81	-.52
Real: RMSE: 2.13	S.D.: 3.33		Separation Index: 1.56		Reliability: .71	
Model: RMSE: 1.72	S.D.: 3.56		Separation Index: 2.07		Reliability: .81	
Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
1st Grade Near Point Accuracy	-.06	.07	1.19	1.15	1.08	.48
1st Grade Far Point Accuracy	.06	.05	.79	-1.20	.52	-.27
Real: RMSE: .75	S.D.: .72		Separation Index: .95		Reliability: .47*	
Model: RMSE: .69	S.D.: .78		Separation Index: 1.12		Reliability: .56	
Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
2nd Grade Near Point Accuracy	-.34	.34	1.33	.77	1.37	.83
2nd Grade Far Point Accuracy	.34	.12	.41	-1.82	.39	-1.07
Real: RMSE: .96	S.D.: 1.18		Separation Index: 1.23		Reliability: .60	
Model: RMSE: .86	S.D.: 1.26		Separation Index: 1.48		Reliability: .69	
Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
3rd Grade Near Point Accuracy	-.31	.35	1.41	.99	1.50	1.08
3rd Grade Far Point Accuracy	.31	.15	.20	-1.46	.20	-1.01
Real: RMSE: .63	S.D.: .32		Separation Index: .50		Reliability: .20*	
Model: RMSE: .58	S.D.: .40		Separation Index: .69		Reliability: .32	

Table 13 (continued)

Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
1st Grade Near Point Fluency	-.22	.09	1.01	.11	.91	-.38
1st Grade Far Point Fluency	.22	.09	1.03	.21	.81	-.34
Real: RMSE: .82	S.D.: .99		Separation Index: 1.20		Reliability: .59	
Model: RMSE: .68	S.D.: 1.10		Separation Index: 1.62		Reliability: .72	

Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
2nd Grade Near Point Fluency	-.4	.09	.97	-.09	.96	-.11
2nd Grade Far Point Fluency	.4	.09	1.06	.34	1.04	.24
Real: RMSE: 1.26	S.D.: 2.06		Separation Index: 1.64		Reliability: .73	
Model: RMSE: 1.06	S.D.: 2.17		Separation Index: 2.06		Reliability: .81	

Grade/Scale	Measure	SE	Infit MNSQ	ZSTD	Outfit MNSQ	ZSTD
3rd Grade Near Point Fluency	.18	.15	1.04	.24	1.06	.37
3rd Grade Far Point Fluency	-.18	.14	.91	-.33	1.07	.36
Real: RMSE: .87	S.D.: 1.09		Separation Index: 1.25		Reliability: .61	
Model: RMSE: .74	S.D.: 1.19		Separation Index: 1.61		Reliability: .72	

*Indicates low separation index and reliability.

Note: The "**model"** person reliability (including measures for extreme scores) is an upper bound to this value when persons are ordered by measures. The "**real"** person reliability (including measures for extreme scores) is a lower bound to this value when persons are ordered by measures.

Rating Scale Functioning

Research question 4. *Is the rating scale of TeleWrite optimized so that it sensitively measures varying levels of handwriting skills, and therefore indicates rating scale responsiveness?*

Hypothesis. The rating scale would monotonically increase in average measures and step calibration of at least 0.5 logit values, indicating the ability to sensitively measure varying levels of handwriting skills which indicates good rating scale responsiveness.

The Partial Credit Model (PCM; Masters, 1982) was used to perform analysis of Rasch step calibrations. The PCM is a unidimensional model for the analysis of responses recorded in two or more ordered categories and contains only two sets of parameters: one for persons and one for items (Wright, 1996b). The Andrich thresholds (step values) are parameters of the PCM. They are the points on the latent variable where adjacent categories of the item are equally probable, that is, one structure per item (Wright, 1996b). To test this hypothesis, this study used the PCM to measure the different component factors (*handwriting rate*, *handwriting accuracy,* and *handwriting fluency*) that comprise the latent variable of handwriting skills. As stated in Chapter III, for test items to be calibrated with more stringency, a sample size of 400 observations is recommended to produce errors below 0.04 logits (Custer, 2015). This study had 148 participants, below the recommended sample size for stringent rating scale analysis using the PCM. The results reported must be interpreted with caution and as preliminary data.

The items of TeleWrite were calibrated in terms of the extent to which the participants' responses on the items corresponded with the item difficulty of the scale. According to Linacre (2002), the Rasch-Andrich Threshold, also known as step-calibration or step measure, when negative step-calibrations occur, may indicate disordering in the rating scale and occurs when some categories are not observed frequently. The Winsteps® analysis generated a report on the performance of the three TeleWrite scales (*rate*, *accuracy,* and *fluency*) from near point (NP) and far point (FP) distances for Grades 1, 2, and 3.

The results of the rating scale function analysis are presented in Table 14 (summary of rating scale analysis), Table 15 (*rate*), Table 16 (*accuracy*), and Table 17

(*fluency*). A summary of analyses for all rating scales is shown in Table 14. All categories in each of the three scales (*handwriting rate, handwriting accuracy, handwriting fluency*) had less than 10 observations due to the low sample size. The number of working categories ranged from 3-9 out of a 10-point scale, with the *accuracy* scale containing the lowest number of working categories and the lowest average measures. All categories in all scales had acceptable outfit MNSQ values, except for *accuracy* scales for Grades 1 and 3 which had outfit MNSQ values > 2.0 logits. Grade 1 near point (category 10) and Grade 3 near and far point (both category 9) had outfit mean square values greater than 2.0. All scales showed one category with disordered advancement of average measures, especially the *accuracy* scale which had disordered measures in every category. Lastly, the step calibrations for several categories did not advance by at least 0.5 logits. Table 15 is a summary of *rate* step-calibration values indicating several disordered items and inconsistent monotonical increases of +/-.05 logits. Similarly, the *accuracy* scales (Table 16) revealed few functional categories, causing the negative step-calibrations and narrow ranges of step calibrations for several categories that did not advance by at least 0.5 logits, especially in the FP dimension. The *fluency* scales (Table 17) appeared to have the most stable calibrations in all categories and for Grades 1-3. Although the categories did not uniformly increase by 0.5 logits in all scales, there was an orderly monotonical increase in all *fluency* categories. As a composite score, the *fluency* scale appeared to be best organized to reliably separate the categories of handwriting skills.

Table 14

Summary of the TeleWrite Rating Scale Analysis Based on Linacre's (2002) Guidelines

Guidelines / Scale/Grade		At Least 10 Observations per Category	Monotonic Increase in Category Measures	Outfit MNSQ <2.0 logits	Step Calibrations Increase by at Least 0.5 logits for 10-item Scale
Rate	**Grade 1**				
	NP	6 observations/category	Yes	Yes	Yes
	FP	5 observations/category	Yes	Yes	Yes
	Grade 2				
	NP	7 observations/category	Yes	Yes	Yes
	FP	5 observations/category	Yes	Yes	Yes
	Grade 3				
	NP	9 observations/category	Category 8 Disordered	Yes	No
	FP	6 observations/category	Category 6 Disordered	Yes	No
Accuracy	**Grade 1**				
	NP	5 observations/category	Category 3 Disordered	No. Category 10 is > 2.0	No
	FP	4 observations/category	Category 8 Disordered	Yes	No
	Grade 2				
	NP	8 observation/category	Category 8 Disordered	Yes	No
	FP	3 observations/category	Category 8 Disordered	Yes	Yes
	Grade 3				
	NP	9 observations/category	Categories 6 & 9 Disordered	No. Category 9 is >2.0	No
	FP	5 observations/category	Category 10 Disordered	No. Category 9 is >2.0	Yes

Table 14 (continued)

Guidelines / Scale/Grade		At Least 10 Observations per Category	Monotonic Increase in Category Measures	Outfit MNSQ <2.0 logits	Step Calibrations Increase by at Least 0.5 logits for 10-item Scale
Fluency	**Grade 1**				
	NP	7 observations/category	Category 2 Disordered	Yes	Yes
	FP	8 observations/category	Category 9 Disordered	Yes	Yes
	Grade 2				
	NP	7 observations/category	Yes	Yes	Yes
	FP	8 observations/category	Category 8 Disordered	Yes	Yes
	Grade 3				
	NP	9 observations/category	Category 5 Disordered	Yes	Yes
	FP	9 observations/category	Categories 5 & 9 Disordered	Yes	No

NP= Near Point FP=Far Point

Table 15

Rating Scale Functioning of Rate Items for 1st, 2nd, and 3rd Graders

Category	Counts Used	Average Measure	Outfit MNSQ	Step Calibration	SE	Category	Counts Used	Average Measure	Outfit MNSQ	Step Calibration	SE
First Grade Rate Near Point						**First Grade Rate Far Point**					
1	2	.00	.00	NONE	.00	1	2	.00	.00	NONE	.00
2	2	.00	.00	NONE	.00	2	5	-5.39	1.20	NONE	.00
3	5	-6.66	.50	NONE	.00	3	11	-3.32	.72	-5.90	.92
4	9	-3.69	1.18	-7.44	1.16	4	7	-.47	.80	-.59	.43
5	11	-.27	1.02	-2.49	.67	5	13	2.74	1.03	1.18	.34
6	9	2.31	.80	1.16	.57	6	12	4.15	1.32	5.31	.36
7	3	4.15	.74	4.04	.72	7	3	.00	.00	.00	.44
8	4	5.40	.44	4.74	.93	8	5	.00	.00	.00	.46
9	2	.00	.00	NONE	.00	9	0	.00	.00	.00	.60
10	1	.00	.00	NONE	.00	10	0	.00	.00	.00	.79
Second Grade Rate Near Point						**Second Grade Rate Far Point**					
1	4	.00	.00	NONE	.00	1	0	.00	.00	NONE	.00
2	7	.00	.00	NONE	.00	2	8	.00	.00	NONE	.00
3	6	-3.73	.94	.00	.00	3	8	-3.38	.90	NONE	.00
4	7	-2.83	1.10	-4.78	.83	4	11	-2.12	.81	-3.31	.59
5	11	-.1.59	.94	-2.86	.52	5	9	-.64	1.36	-1.05	.50
6	10	.15	.82	.88	.51	6	6	1.11	1.04	.79	.62
7	0	.00	.00	NULL	NULL	7	3	.71	.91	3.57	.95
8	2	3.12	.44	4.65	1.12	8	1	3.12	.00	NONE	.00
9	3	3.91	.61	3.87	1.26	9	0	.00	.00	NONE	.00
10	0	.00	.00	.00	.00	10	0	.00	.00	NONE	.00
Third Grade Rate Near Point						**Third Grade Rate Far Point**					
1	4	-1.27	.94	NONE	.00	1	3	-.96	1.12	NONE	.00
2	4	-.77	.48	-1.07	.63	2	5	-.50	1.51	-1.33	.67
3	6	-.54	.79	-1.07	.50	3	16	-.07	1.30	-1.51	.44
4	8	-.04	.81	-.61	.43	4	17	.10	1.27	.04	.33
5	6	.16	.27	.26	.40	5	4	.98	.64	1.94	.44
6	4	.19	.90	.61	.41	6	4	.25*	1.95	.87	.58
7	5	.74	.17	.19	.43	7	0	.00	.00	.00	.00
8	4	.73*	.69	.83	.48	8	0	.00	.00	.00	.00
9	4	1.40	.48	.85	.60	9	0	.00	.00	.00	.00
10	0	.00	.00	.00	.00	10	0	.00	.00	.00	.00

*Disordered measure and no monotonic scale increase.

Table 16

Rating Scale Functioning of Accuracy Items for 1st, 2nd, and 3rd Graders

Category	Counts Used	Average Measure	Outfit MNSQ	Step Calibration	SE	Category	Counts Used	Average Measure	Outfit MNSQ	Step Calibration	SE
First Grade Accuracy Near Point						**First Grade Accuracy Far Point**					
1	0	.00	.00	NULL	.00	1	0	.00	.00	NULL	.00
2	0	-.03	.00	NULL	.41	2	0	.00	.00	NULL	.00
3	0	-.74*	.01	.53	.37	3	0	.00	.00	NULL	.00
4	0	-.48	.12	-.23	.35	4	0	.00	.00	NULL	.00
5	1	-.34	.40	-.17	.35	5	0	.00	.00	NULL	.00
6	0	-.25	.61	-.13	.35	6	0	.00	.00	NULL	.00
7	2	-.18	.68	-.10	.37	7	2	-.33	.47	1.27	1.07
8	8	-.12	.70	-.08	.41	8	2	-.62*	.00	-.32	.78
9	11	-.07	.72	-.06	.43	9	6	.54	.76	-.53	.58
10	26	-.03	2.27	.24	.61	10	33	.72	1.02	-.42	.41
Second Grade Accuracy Near Point						**Second Grade Accuracy Far Point**					
1	10	-1.11	.49	NONE	.00	1	0	.00	.00	NULL	.00
2	4	-.40	.20	-2.07	.66	2	0	.00	.00	NULL	.00
3	4	.32	.08	-1.61	.58	3	0	.00	.00	NULL	.00
4	4	.59	.16	-1.17	.57	4	0	.00	.00	NULL	.00
5	4	1.54	.01	-.67	.60	5	0	.00	.00	NULL	.00
6	0	1.75	.97	.08	.68	6	0	.00	.00	NULL	.00
7	4	4.35	.00	1.52	.80	7	1	.32	4.69	NULL	.00
8	4	4.35*	.95	3.92	1.10	8	0	.00	.00	NULL	.00
9	0	.00	.00	NONE	.00	9	10	-.14*	7.14	-1.11	1.07
10	0	.00	.00	NONE	.00	10	33	1.28	1.09	1.11	.44
Third Grade Accuracy Near Point						**Third Grade Accuracy Far Point**					
1	9	-.27	.70	NONE	.00	1	2	-.38	.24	NONE	.00
2	4	-.18	.51	-.53	.56	2	0	0	.00	NULL	.00
3	4	-.05	.23	-.48	.45	3	0	0	.00	NULL	.00
4	4	.14	.05	-.40	.43	4	0	0	.00	NULL	.00
5	4	.36	.01	-.29	.43	5	0	0	.00	NULL	.00
6	4	.34*	.45	.09	.46	6	0	0	.00	NULL	.00
7	5	.73	.83	.01	.50	7	1	-.05	.46	1.88	.80
8	4	1.66	1.20	.84	.54	8	1	.14	.83	.33	.66
9	7	1.16*	1.44	.93	.62	9	4	1.21	7.00	-.93	.58
10	0	.00	.00	.00	.00	10	37	.40*	1.03	-1.29	.43

*Disordered measure and no monotonic scale increase.

Table 17

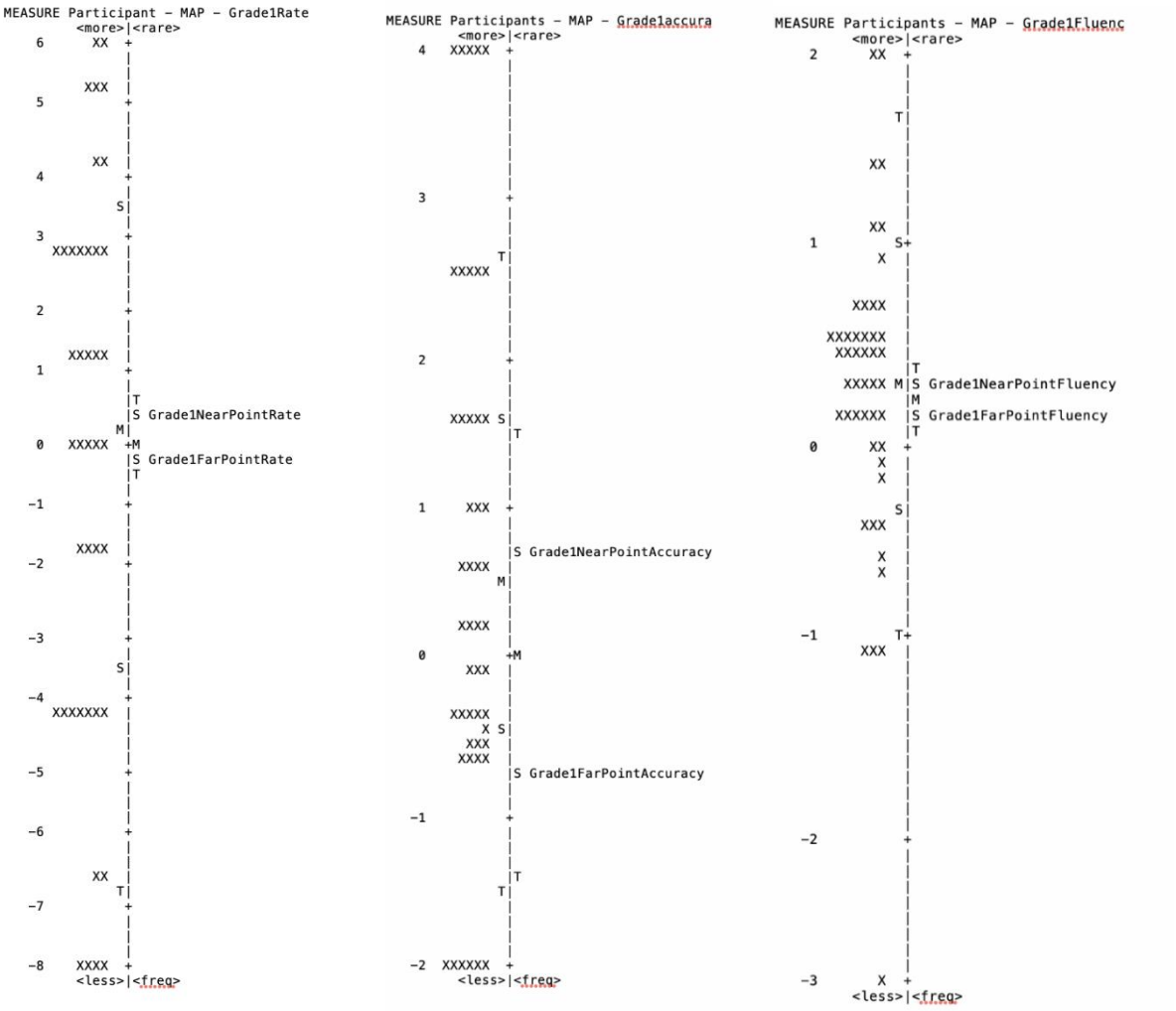
Rating Scale Functioning of Fluency Items for 1st, 2nd, and 3rd Graders

Category	Counts Used	Average Measure	Outfit MNSQ	Step Calibration	SE	Category	Counts Used	Average Measure	Outfit MNSQ	Step Calibration	SE
First Grade Fluency Near Point						**First Grade Fluency Far Point**					
1	5	NONE	.00	NONE	.00	1	4	-.34	4.89	NONE	.00
2	2	NONE	.00	NONE	.00	2	0	.00	.00	NULL	.00
3	3	-2.46	.48	NONE	.00	3	6	-2.01*	.31	-4.28	.86
4	5	-1.84	.17	-2.77	.80	4	8	-.94	.28	-1.13	.51
5	8	-.69	.34	-1.75	.52	5	8	-.36	.47	-.31	.44
6	11	-.04	.72	-1.01	.43	6	5	.60	.20	.70	.48
7	4	.77	.04	.92	.49	7	5	.95	.30	.84	.53
8	4	1.55	.54	.55	.55	8	4	1.96	.11	1.64	.59
9	5	1.37*	1.87	.90	.59	9	2	1.76*	1.04	2.55	.81
10	1	-.11*	1.30	3.15	1.07	10	0	.00	.00	NULL	.00
Second Grade Fluency Near Point						**Second Grade Fluency Far Point**					
1	7	NONE	.00	NONE	.00	1	NONE	NONE	.00	NONE	.00
2	5	NONE	.00	NONE	.00	2	6	NONE	.00	NONE	.00
3	6	-4.91	.70	NONE	.00	3	2	NONE	.00	NONE	.00
4	5	-3.73	1.60	-4.81	.84	4	8	-3.65	1.36	NONE	.00
5	11	-2.11	.45	-3.26	.62	5	10	-2.69	1.10	-3.92	.60
6	8	-.72	.70	-.59	.51	6	8	-.70	.58	-1.72	.51
7	5	1.21	.97	1.13	.66	7	6	1.57	.38	-.15	.63
8	2	2.98	1.02	3.22	.89	8	2	1.10*	1.77	2.40	.87
9	1	3.37	.95	4.32	1.28	9	1	4.34	.33	3.39	1.28
10	0	NONE	.00	.00	.00	10	NONE	.00	.00	.00	.00
Third Grade Fluency Near Point						**Third Grade Fluency Far Point**					
1	4	-.96	.89	NONE	.00	1	0	0	0	NONE	.00
2	4	-.45	.56	-.69	.62	2	2	-.93	1.03	NONE	.00
3	4	-.40	1.55	-.40	.49	3	3	-.76	1.06	-1.22	.81
4	4	.04	1.70	-.20	.43	4	10	-.10	1.29	-1.65	.54
5	6	-.01*	.63	-.45	.40	5	10	-.27*	.50	-.20	.36
6	7	.04	.90	-.04	.47	6	10	.02	1.18	-.01	.33
7	7	.45	.39	.26	.37	7	7	.36	1.03	.51	.37
8	5	.48	.97	.75	.42	8	1	.57	.21	2.26	.47
9	4	.63	.89	.93	.56	9	4	.46*	1.25	-.94	.49
10	0	.00	.00	NONE	.00	10	2	.58	1.08	1.25	.76

*Disordered measure and no monotonic scale increase.

Figure 3

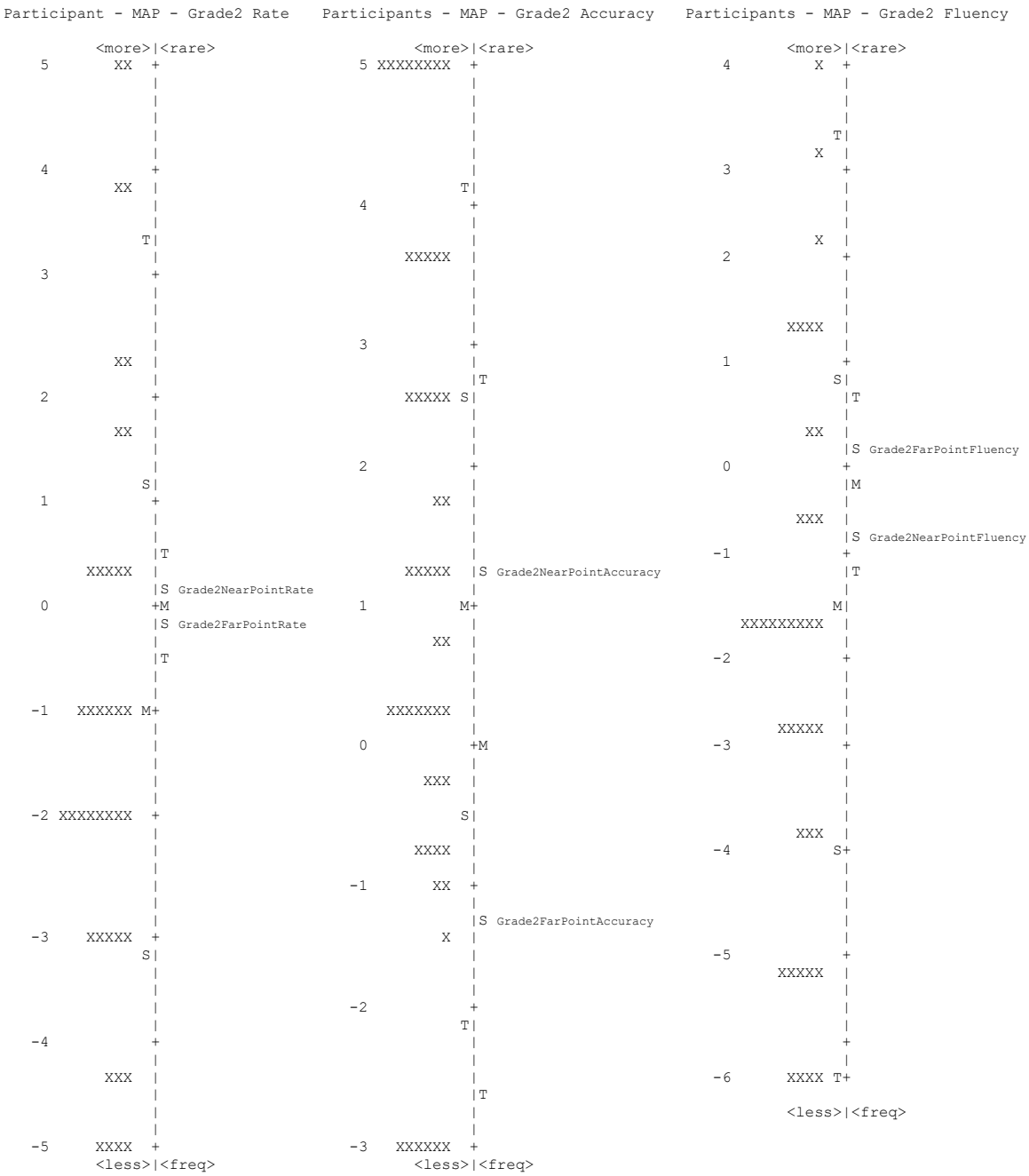
Vertical Rulers of 1st Grade TeleWrite Items



Note. Numbers on the left side indicate the measures of person ability and item difficulty in logits. Each X on the left side of the map indicates 1 participant; items are located on the right side of the map. The M to the left of the vertical line is the mean of the person abilities, and the M to the right of the vertical line is the mean of the item difficulties. S indicates 1 standard deviation and T indicates 2 standard deviations for item difficulty and person ability, respectively.

Figure 4

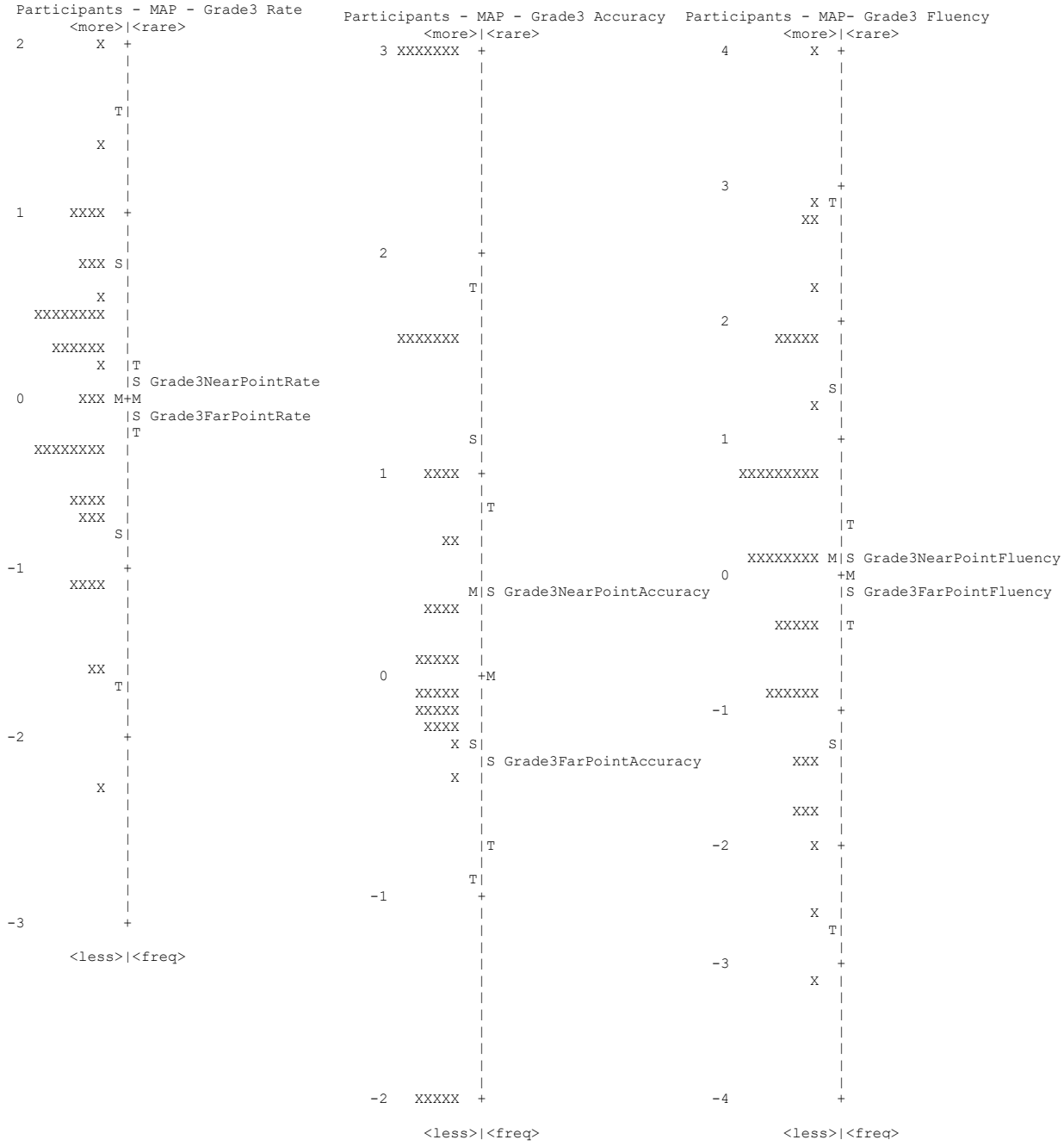
Vertical Rulers of 2nd Grade TeleWrite Items



Note. Numbers on the left side indicate the measures of person ability and item difficulty in logits. Each X on the left side of the map indicates 1 participant; items are located on the right side of the map. The M to the left of the vertical line is the mean of the person abilities, and the M to the right of the vertical line is the mean of the item difficulties. S indicates 1 standard deviation and T indicates 2 standard deviations for item difficulty and person ability, respectively.

Figure 5

Vertical Rulers of 3rd Grade TeleWrite Items



Note. Numbers on the left side indicate the measures of person ability and item difficulty in logits. Each X on the left side of the map indicates 1 participant; items are located on the right side of the map. The M to the left of the vertical line is the mean of the person abilities, and the M to the right of the vertical line is the mean of the item difficulties. S indicates 1 standard deviation and T indicates 2 standard deviations for item difficulty and person ability, respectively.

Chapter VII

DISCUSSION

The findings of four preliminary studies examined the development and validation of a new telehealth-based handwriting assessment. All preliminary studies supported the continued development of the TeleWrite assessment by advancing to construct validation of the tool using Rasch methods.

Interrater Reliability

This preliminary study demonstrated that TeleWrite has an excellent interrater reliability of 0.92, at a 95% confidence interval and a Cronbach's α measure for internal consistency of 0.95. The results indicated that TeleWrite has excellent interrater reliability in assessing the handwriting skills of children in Grades 1-3. The preliminary results also indicated that TeleWrite can be a reliable handwriting assessment when administered with children between the ages of 6 and 8 via telehealth. The statistical results suggested a high correlation of interrater reliability for total scores, capital letters, lower-case letters, and numerals. However, additional psychometric testing is warranted for later phases of development.

Clinical Utility

A clinical utility study of 55 pediatric occupational therapists from 12 countries and 22 U.S. states rated TeleWrite as an instrument that is easy to use, useful in clinical practice, and accurate in assessing handwriting skills. The results of the survey were analyzed, and four main themes were identified: (a) TeleWrite is easy to use and has an easy-to-follow format; (b) TeleWrite is a potentially useful tool in practice; (c) need for clearer examples and clearer instructions; and (d) potential to include other observable variables. A total of 64% of survey respondents rated that this tool "quite/extremely accurately" measures handwriting performance. However, clinicians who received a script for administration and instructions found areas of ambiguity. Further development of an administration and scoring manual is needed.

Content Validity

Content validity is determined by the relationship between the definition of a construct (i.e., handwriting skills) and the items designed to measure it (Dixon & Johnston, 2019). A content validity study of nine handwriting experts concluded that all 10 subtests of the TeleWrite tool are considered essential. Further, to achieve an excellent content validity rating, the item content validity or I-CVI should be 0.78 or higher for three or more experts, and the scale content validity index S-CVI should be 0.90 or higher (Polit et al., 2007). The TeleWrite tool surpassed these thresholds at I-CVI of 0.88 and S-CVI of 0.94. While CVI examines the relevance of test items in relation to a single construct (Dixon & Johnston, 2019), it does not quantify the extent to which the measure is distinct and not altered by other constructs. In response, the researcher decided to examine the construct validity of the tool.

Construct Validity, Reliability, and Rating Scale Functioning

In the Rasch Analysis study, the researcher determined that, overall, the TeleWrite scales fit the unidimensional Rasch model of measurement and established construct and internal validity when analyzed as three separate sub-latent constructs or domains. This research determined that TeleWrite is comprised of three distinct and unidimensional scales: *handwriting rate, handwriting accuracy,* and *handwriting fluency.* These three variables have been operationally defined in this manuscript as *writing rate,* or the number of letters in the writing prompt copied by the student in minutes*; writing accuracy,* or the total number of letters in the writing prompt that are recognizable within context; and *writing fluency,* which is a derived score calculated by dividing the total number of letters written correctly from the *"accuracy"* score by the *"rate"* score in minutes. Students who achieve writing fluency (i.e., accuracy and speed of handwritten letters) may be able to expend more cognitive and motor effort on other higher writing processes such as writing from memory or dictation (Staats et al., 2019).

Capturing and quantifying the facets of *handwriting fluency* is a unique contribution of this study. Writing fluency is considered the most accurate index of automation (Kim et al., 2014; Palmis et al., 2017). In the present study, automaticity or fluency of handwriting was found to be a reliable measure of handwriting. The *fluency* scales of third graders showed most scores above the mean, possibly indicating that the third graders were beginning to achieve writing automaticity. In addition, the present study corroborated the results of previous studies demonstrating that the speed of manual writing increases at each grade level and improvement in speed may be explained by better automation of handwriting skills (Graham et al., 1998; Palmis et al., 2017).

Most participants in all grades endorsed distance (far point) writing more easily, as indicated by less item difficulty (Rasch average measures). These results are supported by previous research that found that the average speed was 5.4 letters per minute faster for first graders and 7.2 letters per minute faster for second graders on far copy tasks, as compared to near point or close copy tasks (Couture et al., 2016). Far point or distance copying requires the student to read and hold information in short-term memory and transcribe to paper. Near point (from a book, etc.) may or may not be easier because the distance between source and copy is shorter so the transfer can be "quicker" (Tseng & Chow, 2000). In near point copying, because the ability to hold small amounts of information in the brain is challenged, the student takes smaller chunks of information (i.e., a few letters) or words rather than a whole sentence, thus requiring more visual glances and slower transcription speed onto the paper (Tseng & Chow, 2000).

Rasch analyses provided insights into how to modify the tool to demonstrate better fit with the Rasch model. Given that handwriting is a multidimensional construct in theory (Amudson, 1995; Reisman, 1999; Stievano et al., 2016), separate Rasch calibrations were determined to be the most appropriate way to analyze the unidimensionality of the tool.

The person separation reliability is equivalent to CTT's Cronbach's α; it provides an estimate of how well a tool can differentiate persons on the measured variable (Bond & Fox, 2015). A person separation index of 1.5 or a person reliability coefficient of 0.70 represents an acceptable level of separation and is considered the minimum required to divide the sample into two distinct strata (i.e., low and high ability), although the preferred values for reliability are 0.80 and 2.0 for separation ratio (Fisher, 1992).

The results of this study for Grades 1-3 indicated an acceptable separation reliability or correlation coefficient for *rate* (0.68-0.76) and *fluency* (0.61-0.73). However, the *accuracy* scales for Grades 1-3 were in the poor-fair range (0.20-0.60), indicating the ability of the scale to only detect a narrow range of participants' performance and decreased test sensitivity (Boone & Noltenmeyer, 2017). These values, however, need further study as the sample was not large enough to confirm the item difficulty hierarchy. Because TeleWrite was divided into three separate scales after a unidimensionality analysis, more participants are needed per scale to better distinguish between the high and low ability levels of the participants' performance. Person separation measures are influenced by factors such as length of the scale, number of categories per item, and match between the items and the ability of the respondents (i.e., sample item targeting) (Linacre, 2020).

To further analyze and discuss the data obtained in the present study, the results were compared to the psychometric properties of four normed or criterion-referenced handwriting assessments currently widely used in OT clinical practice. The currently available handwriting assessments were designed for in-person administration and have not been validated for telehealth administration. These assessment tools were validated 15-30 years ago using classical test theory (CTT). Because, to date, no other handwriting assessment has been Rasch analyzed, only reliability study comparisons can be made between TeleWrite and existing handwriting assessment instruments. TeleWrite was Rasch analyzed for separate calibrations per scale and with no composite scores, as normally reported in CTT. This feature allows clinicians to identify and report separate scores for *handwriting rate*, *accuracy*, and *fluency* for first, second, and third graders.

The Test of Handwriting Skills-Revised (THS-R) (Milone, 2007) is the norm-referenced, updated version of the original Test of Handwriting Skills (Gardner, 1998). The internal consistency reliability coefficients for Grades 1-3 ranged from poor to good (0.57 to 0.87) for individual subtests, and the test-retest reliability of the THS-R was highly variable as it ranged from 0.37 to 0.82 (Milone, 2007). It is important to note that the reliability coefficient scores denote "overall" total scores for all age groups from 6-18+ years. The Children's Handwriting Evaluation Scale for Manuscript Writing (CHES-M) (Phelps & Stemple, 1987) is a norm-referenced tool with scoring standards based on children in Grades 1 and 2. Correlations for interrater reliability for two raters were in the moderate range (0.65-0.81); for three raters, it was in the excellent range (0.85-0.93) (Phelps & Stemple, 1987). Validity studies have not been reported. As Feder and Majnemer (2003) stated, the scoring system is unclear and the criteria for the categories are not well defined.

The Evaluation Tool of Children's Handwriting-Manuscript (ETCH-M) (Amundson, 1995) is a criterion-referenced assessment. The reported interrater reliability with intraclass correlation coefficients (ICCs) ranged from 0.42 to 0.84 for individual manuscript items. There are no normative data, and information received from the ETCH is qualitative because of the lack of normative samples (Feder & Majnemer, 2003). The Minnesota Handwriting Assessment (MHA) is a norm-referenced tool developed by Reisman (1999). The interrater reliability correlation coefficients for total accuracy scores are 0.72, with a between-school range of $r = 0.58$ to 0.94 (Feder & Majnemer, 2003). The MHA is only standardized for use with children in Grades 1 and 2 (Reisman, 1999). This assessment has a long and detailed scoring, and some components are subject to

interpretation (Roston et al., 2008). The manual described an informal content validation study, but to date, no formal internal structure constructs of validity have been published.

The moderate reliability values of the TeleWrite obtained through Rasch methods represent acceptable evidence of reliability coefficients comparable to the most often utilized handwriting assessments (THS-R, CHES-M, MHA, and ETCH). The Rasch methods used, and the results of this study provide confirmation of the complexity involved in capturing all facets of the construct of handwriting skills. Current handwriting assessments do not quantify specific constructs of handwriting and do not report specificity or sensitivity to change (as reported by the Rasch separation statistic), limiting their usefulness to therapists (Feder & Majnemer, 2003). The lack of valid and reliable handwriting evaluation tools and the complexity in the scoring of existing tools limit the usage of standardized assessments in the evaluation of handwriting skills. Handwriting assessments are not widely used by occupational therapists in schools or clinical settings (Feder, Majnemer, & Synnes, 2000).

Handwriting is a complex perceptual-motor skill that is dependent on the maturation and integration of cognitive, perceptual, and motor skills (Berninger & Fuller, 1992). The evaluation of children's handwriting skills remains a challenge for pediatric therapists. The complexity of handwriting requires the integration of information from various body systems, leading to variability in individual performance from day to day, especially in beginner writers. Such variability may impact the natural subjectivity of evaluating handwriting and, correspondingly, designing a handwriting assessment with strong validity and reliability is complex due to the multifaceted characteristics of writing skills. The TeleWrite assessment was developed to measure the constructs of handwriting

reflected in theory and clinical occupational therapy pediatric practice. A recent systematic review of the literature (Grajo, Candler, & Sarafian, 2020) provided strong evidence in support of addressing important components of therapeutic handwriting practice to include pencil grasp, letter formation, line placement, letter size, letter spacing, and legibility. The TeleWrite assessments include all these components for assessment as tenets of handwriting skills.

TeleWrite seems to be able to assess average low and average high abilities of the participants rather than extremely low and extremely high abilities. Although a person separation of 3 strata or levels is recommended, the separation indices in this study indicated that the TeleWrite items were enough to discriminate the sample in two groups with low average and high average handwriting ability. The separation reliability of the *accuracy* scales fell below established benchmarks. These findings may be explained by a small sample size which would increase the standard errors (SE) and therefore reduce reliability (Daher, Ahmad, Winn, & Selamat, 2015). One way to strengthen the findings is to increase the sample of participants. Another solution is to decrease the number of categories of the *accuracy* scale from 10 categories. Daher et al. (2015) found improved person and item reliability and better fit statistics by adjusting the number of rating scale categories and rerunning iterations. Lastly, the researcher may reconfigure *accuracy* scores based on the five dimensions of handwriting accuracy identified by Reisman (1999). Scores may be assigned from 0 to 5 points, assigning one point per criterion of legibility, form, size, spacing, and alignment, rather than the current 0-3 criteria (0: Letter is not recognizable out of context; 1: Letter is somewhat recognizable in context; 2: Letter clearly

resembles the model). Broader assessment criteria for *accuracy* may decrease subjectivity when scoring accuracy items during evaluation.

Implications for Occupational Therapy Practice

This study has many implications for occupational therapy practice. The American Occupational Therapy Foundation (AOTF, 2015) has identified telehealth and health information communications technologies as crucial research priorities to improve healthcare access and quality and to meet the needs of medically underserved areas and populations. Currently, no pediatric handwriting assessments have been validated for telehealth use and the TeleWrite assessment would fill an unmet service need. The benefit for OT practice is the ability to access underserved populations by limiting the distance between clinicians. This dissertation included children and families from several different U.S. states and international locations who connected with the researcher in real time through use of their smartphone, computer, or tablets. Their participation in the present study may not have been possible without the use of telehealth technology.

The present research broadens the use and validation of pediatric assessments for telehealth practice and extends our scope of practice. TeleWrite is a telehealth-based handwriting assessment that can be useful in detecting handwriting challenges in elementary school children and may be used as a baseline measure to assess the handwriting services provided by occupational therapists in a variety of settings, including schools, home health, and outpatient clinics.

Limitations and Future Research

The generalizability of this study is limited because a convenience sample was utilized mostly from the northeastern United States, which excluded certain demographic regions and associated socioeconomic backgrounds. This omission may overlook the increasing racial and ethnic diversity in the United States. In the current study, differential item functioning (DIF) by gender, ethnicity or urban vs suburban settings was not explored. These variables may influence the participants' ability to endorse handwriting test items even when the participants are in the same grade level (1-3). Gender differences do exist in developing handwriting abilities and handwriting dysfunction. Handwriting studies have demonstrated that girls significantly outperformed boys in improving handwriting quality and speed (Berninger & Fuller, 1992; Feder & Majnemer, 2007). It is recommended that future research use item response theory approach to explore DIF among gender, ethnicity, and settings (metropolitan, suburban and rural settings).

TeleWrite was Rasch analyzed to optimize the psychometric properties of the instrument. However, following the Rasch partial credit model, a larger sample is necessary to obtain improved calibration, reliability, and validity measures. The handwriting *rate* results may need further investigation due to the wide variability of speed in all grades. Data collection for this study occurred 6 months into the COVID-19 pandemic (September-December 2020), when most children were receiving virtual education and completing digital written assignments. This atypical mode of instruction, beyond the researcher's control, may have influenced the students' *rate* of writing, possibly due to decreased attention and motivation for engaging in writing tasks. It is also

important to consider the timing of the assessment, as the *writing rate* (speed) may vary throughout the school year. Since the data were collected in the first quarter of the school year, handwriting skills may have been less developed than in the middle of the school year. The unprecedented conditions under which handwriting *rate* data were collected warrant further research under more typical school conditions. The TeleWrite also has the potential to be used for in-person administration. To explore this, future research is needed to validate the tool using in-person administration.

In addition, to examine patterns of intercorrelations among the TeleWrite measures in correlational results, convergent and discriminant validity, subtypes of construct validity, may be analyzed through correlation coefficients. The purpose is to establish if similar constructs between TeleWrite and other handwriting or visual perceptual tools correspond with one another. Qualitative methods including focus groups and clinician interviews may be conducted to improve the range and intelligibility of TeleWrite items.

Conclusion

In view of the results of the present study, TeleWrite appears to have promising potentials to detect and discriminate for differences in handwriting abilities among elementary school children. The study provided preliminary evidence of the validity and reliability of TeleWrite as a clinically useful tool. The findings are encouraging, especially considering the challenges of defining the constructs of a complex task such as handwriting as well as developing good measurement tools to capture these constructs. There is consensus in the OT profession that handwriting is a multifactorial construct that requires the interaction of cognitive and motor processes (Feder & Majnemer, 2007).

This exploratory Rasch analysis demonstrated that TeleWrite had acceptable psychometric properties as a unidimensional measure of children's handwriting skills in Grades 1-3. The instrument's three distinct handwriting constructs of *rate, accuracy,* and *fluency* cover a wide spectrum of handwriting challenges found in elementary school children. Future studies are needed with additional participants to further improve the tool. However, the findings of this study suggested that practitioners can be confident in their interpretation of TeleWrite as an instrument for use in evaluating children's handwriting skills. Finally, the findings of the current research contribute to the present literature as the first handwriting assessment specifically designed and validated for telehealth use to assess *handwriting rate, handwriting accuracy and handwriting fluency*, all pertinent variables of handwriting associated with handwriting dysfunction.

REFERENCES

Ailliet, L., Knol, D. L., Rubinstein, S. M., de Vet, H. C. W., van Tulder, M. W., & Terwee, C. B. (2013). Definition of the construct to be measured is a prerequisite for the assessment of validity. *Journal of Clinical Epidemiology, 66*(7), 775-782.

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Waveland Press.

American Occupational Therapy Association (AOTA). (2013). Telerehabilitation. *American Journal of Occupational Therapy, 64*(Suppl.), 92-102.

American Occupational Therapy Association (AOTA). (2018). *Occupational therapy and telehealth state statutes, regulations, and regulatory board statements.* https://www.aota.org/~/media/Corporate/Files/Secure/Advocacy/State/Telehealth-State-StatutesRegulations-Regulatory-Board-Statements.pdf.

American Occupational Therapy Association (AOTA). (2020). *Information pertaining to occupational therapy in the era of Coronavirus (COVID-19)*. https://www.aota.org/Practice/Health-Wellness/COVID19.aspx

American Occupational Therapy Foundation (AOTF). (2015, October). *Research priorities to advance full participation in meaningful life activities.* https://www.aotf.org/About-AOTF/Research-Priorities.

American Psychological Association (APA). (2020, May 1). *Guidance on psychological tele-assessment during the COVID-19 crisis.* https://www.apaservices.org/practice/reimbursement/health-codes/testing/tele-assessment-covid-19

Amundson, S. J. (1995). *Evaluation tool of children's handwriting*. Homer, OTKIDS.

Andrich, D., & Marais, I. (2019). A *course in Rasch measurement theory: Measuring in the educational, social, and health sciences*. Springer.

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement Transactions, 22,* 1145-1146. https://www.rasch.org/rmt/rmt221a.htm

Balog, C. R. (2016). A validation of the efficacy of descriptive instrumental collective case study research methodology for examining pilot cognitive functioning. *International Journal of Aviation, Aeronautics, and Aerospace, 3*(4).

Behl, D., Blaiser, K., Cook, G., Barrett, T., Callow-Heusser, C., Brooks, B., Dawson, P., Quigley, S., & White, K., (2017). A multisite study evaluating the benefits of early intervention via telepractice. *Infants and Young Children, 30*(2), 147-161.

Benham, S., & Gibbs, V. (2017). Exploration of the effects of telerehabilitation in a school-based setting for at-risk youth. *International Journal of Telerehabilitation, 9,* 39-46.

Berninger, V. W., Abbott, R. D., Abbott, S. P., Graham, S., & Richards, T. (2002). Writing and reading: Connections between language by hand and language by eye. *Journal of Learning Disabilities, 35*(1), 39-56. doi:http://dx.doi.org.ezp.scranton.edu/10.1177/002221940203500104

Berninger, V. W., & Fuller, F. (1992). Gender differences in orthographic, verbal, and compositional fluency: Implications for assessing writing disabilities in primary grade children. *Journal of School Psychology, 30,* 363-382.

Bobak, C., Barr, P. & O'Malley, A. (2018). Estimation of an interrater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. *BMC Medical Research Methodology, 18,* 93. https://doi.org/10.1186/s12874-018-0550-6

Bond, T., & Fox, C. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Erlbaum.

Boone W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE Life Sciences Education, 15*(4), 4. doi:10.1187/cbe.16-04-0148

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education, 4*(1). https://dx.doi.org.ezproxy.cul.columbia.edu/10.1080/2331186X.2017.1416898

Brossard-Racine, M., Mazer, B., Julien, M., & Majnemer, A. (2012). Validating the use of the evaluation tool of children's handwriting-Manuscript to identify handwriting difficulties and detect change in school-age children. *American Journal of Occupational Therapy, 66,* 414-421.

Brown, T. & Unsworth, C. (2009) Construct validity of the Test of Visual-Motor Skills-Revised (TVMS–R): An evaluation using the Rasch Measurement Model, *Scandinavian Journal of Occupational Therapy*, *16:*3, 133-145, DOI: 10.1080/11038120802443662.

Bruininks, R., & Bruininks, B. (2005). *Bruininks-Oseretsky test of motor proficiency* (2nd ed.). NCS Pearson.

Bureau of Labor Statistics, U.S. Department of Labor. (2020). *Occupational outlook handbook*. https://www.bls.gov/ooh/healthcare/occupational-therapists.htm

Cason, J. (2014). Telehealth: A rapidly developing service delivery model for occupational therapy. *International Journal of Telerehabilitation, 6*(1), 29-36.

Cason, J., Behl, D., & Ringwalt, S. (2015). Overview of states' use of telehealth for delivery of early intervention (IDEA Part C) services. *Journal of Telerehabilitation, 4*(2), 39-45.

Castorr, A. H., Thompson, K. O., Ryan, J. W., Phillips, C. Y., Prescott, P. A., & Soeken, K. L. (1990). The process of rater training for observational instruments: Implications for interrater reliability. *Research in Nursing and Health, 13*(5), 311-318. https://doi.org/10.1002/nur.4770130507

Centers for Medicare and Medicaid Services (CMS). (2020, March 30). *List of telehealth services.* CMS.gov. https://www.cms.gov/Medicare/Medicare-General-Information/Telehealth/Telehealth-Codes

Cermak, S. A., & Bissell, J. (2014). Content and construct validity of Here's How I Write (HHIW): A child's self-assessment and goal setting tool. *American Journal of Occupational Therapy, 68*(3), 296-306.

Chang, S., & Yu, N. (2010). Characterization of motor control in handwriting difficulties in children with or without developmental coordination disorder. *Developmental Medicine and Child Neurology*, *52*(3), 244-250.

Collins, W. (2008). *Evidence Topic: Handwriting Assessment* [Critically Appraised Topic]. American Occupational Therapy Association.

Couture, M., Morin, M., Coallier, M., Lavigne, A., Archambault, P., Bolduc, É., ... Jasmin, E. (2016). Handwriting assessment of Franco Québec primary school-age students. *The Canadian Journal of Occupational Therapy, 83*(5), 269-280. https://dx.doi.org.ezproxy.cul.columbia.edu/10.1177/0008417416661573

Criss, M. J. (2013). School-based telerehabilitation in occupational therapy: Using telerehabilitation technologies to promote improvements in student performance. *International Journal of Telerehabilitation, 5*(1), 39-46.

Custer, M. (2015, October). Sample size and item parameter estimation precision when utilizing the one-parameter "Rasch" model. Paper presented at the Mid-Western Educational Research Association Annual Meeting, Evanston, Illinois.

Daher, A. M., Ahmad, S. H., Winn, T., & Selamat, M. I. (2015). Impact of rating scale categories on reliability and fit statistics of the Malay Spiritual Well-Being Scale using Rasch analysis. *The Malaysian Journal of Medical Sciences: MJMS, 22*(3), 48-55.

DeVellis, R. F. (2012). *Scale development. Theory and applications* (3rd ed.). Sage.

Dixon, D., & Johnston, M. (2019). Content validity of measures of theoretical constructs in health psychology: Discriminant content validity is needed. *British Journal of Health Psychology, 24*, 477-484. https://doi-org.ezproxy.cul.columbia.edu/10.1111/bjhp.12373

Donica, D. K., & Holt, S. (2019). Examining validity of the Print Tool compared with Test of Handwriting Skills–Revised. *OTJR: Occupation, Participation and Health, 39*(3), 167-175. https://doi.org/10.1177/1539449218804529

Dunn, W., Brown, C., & McGuigan, A. (1994). The ecology of human performance: A framework for considering the effect of context. *American Journal of Occupational Therapy, 48*(7), 595-607. htpps://doi:10.5014/ajot.48.7.595

Dunn, W., McClain, L. H., Brown, C., & Youngstom, M. J. (1998). *The ecology of human performance*. In M. E. Neistadt & E. G. Crepeau (Eds.), *Willard and Spackman's occupational therapy* (9th ed., pp. 531-534). Lippincott.

Engel, C., Lillie, K., Zurawski, S., & Travers, B. G. (2018). Curriculum-based handwriting programs: A systematic review with effect sizes. *American Journal of Occupational Therapy, 72*(3), 1-8,1A-7A. htpps://doi:10.5014/ajot.2018.027110

Fawcett, A. (2009). *Principles of assessment and outcome measurement for occupational therapists and physiotherapists: Theory, skills and application.* John Wiley & Sons.

Feder, K. P., & Majnemer, A. (2003). Children's handwriting evaluation tools and their psychometric properties. *Physical and Occupational Therapy in Pediatrics, 23*(3), 65-84.

Feder, K. P., & Majnemer, A. (2007), Handwriting development, competency, and intervention. *Developmental Medicine and Child Neurology, 49*, 312-317. https://doi:10.1111/j.1469-8749.2007.00312

Feder, K. P., Majnemer, A., & Synnes, A. (2000). Handwriting: current trends in occupational therapy practice. *Canadian Journal Occupational Therapy, 67*,(3), 197-204.

Ferguson, J., Craig, E. A., & Dounavi, K. (2018). Telehealth as a model for providing behaviour analytic interventions to individuals with autism spectrum disorder: A systematic review. *Journal of Autism and Developmental Disorders*. https://doi:10.1007/s10803-018-3724.

Fisher, W. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transactions, 6,* 238. https://www.rasch.org/rmt/rmt63i.htm

Francher, L. A., Priestley-Hopkins, D. A., & Jeffries, L. M. (2018). Handwriting acquisition and intervention: A systematic review. *Journal of Occupational Therapy, Schools and Early Intervention, 11*(4), 454-473. https://doi.org/10.1080/19411243.2018.1534634

Gardner, M. F. (1998). *THS, test of handwriting skills: Manual*. Psychological and Educational Publications.

Gentile, A. M. (1972). A working model of skill acquisition with application to teaching. *Quest, 17*(1), 3-23.

Gentile, A. M. (1998). Implicit and explicit processes during acquisition of functional skills. *Scandinavian Journal of Occupational Therapy, 5*, 7-16.

Gibbs, V., & Toth-Cohen, S. (2011). Family-centered occupational therapy and telerehabilitation for children with autism spectrum disorders. *Occupational Therapy in Health Care, 25*(4), 298-314.

Graham, S. (1992). Issues in handwriting instruction. *Focus on Exceptional Children, 25*(2), 1-13.

Graham, S., Berninger, V. W., & Weintraub, N. (1998). Development of handwriting speed and legibility in grades 1-9. *Journal of Educational Research, 92*, 42-52.

Grajo, L., Candler, C., Bowyer, P., Schultz, S., & Thomson, J. (2018). The clinical and classroom utility of the inventory of reading occupations: An assessment tool of children's reading participation. *The Open Journal of Occupational Therapy, 6*, 1-10. https://doi.org/10.15453/2168-6408.1440

Grajo, L. C., Candler, C., & Sarafian, A. (2020). Interventions within the scope of occupational therapy to improve children's academic participation: A systematic review. *American Journal of Occupational Therapy, 74*, 7402180030. https://doi.org/10.5014/ajot.2020.039016

Grajo, L. C., Gutman, S. A., Gelb, H., et al. (2019). A Rasch analysis study determining the internal validity and separation reliability of the Inventory of Reading Occupations—Adult: A measure of functional literacy participation. *Occupational Therapy in Health Care, 33*(4), 394-412. https://doi:10.1080/07380577.2019.1678216

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*, 38–47. https://dx.doi.org/ 10.1111/j.1745-3992. 1993.tb00543

Hand, B. N., Darragh, A. R., & Persch, A.C. (2018). Thoroughness and psychometrics of fidelity measures in occupational and physical therapy: A systematic review. *The American Journal of Occupational Therapy, 72*, 1-10. https://doi:10.5014/ajot.2018.025510

Heimerl, S., & Rasch, N. C. (2009). Delivering developmental occupational therapy consultation services through telehealth. *Developmental Disabilities Special Interest Section Quarterly, 32*, 1-4.

Heitzman-Powell, L. S., Buzhardt, J., Rusinko, L. C., & Miller, T. M. (2014). Formative evaluation of an ABA outreach training program for parents of children with autism in remote areas. *Focus on Autism and Other Developmental Disabilities, 29*(1), 23-38.

Henderson, S. E., Sugden, D. A., & Barnett, A. L. (2007). *Movement Assessment Battery for Children-2* (2nd ed.). Pearson Assessment.

Hepburn, S. L., Blakeley-Smith, A., Wolff, B., & Reaven, J. A. (2016). Telehealth delivery of cognitive-behavioral intervention to youth with autism spectrum disorder and anxiety: A study. *Autism, 20*(2), 207-218. https://doi:10.1177/1362361315575164

Hodge, M., Sutherland, R., Jeng, K., Bale, G., Batta, P., Cambridge, A., Detheridge, J., Drevensek, S., Edwards, L., Everett, M., & Silove, N. (2018). Agreement between telehealth and face-to-face assessment of intellectual ability in children with specific learning disorder. *Journal of Telemedicine and Telecare, 25*(7), 431-437.

Ingersoll, B., Shannon, K., Berger, N., Pickard, K., & Holtz, B. (2017). Self-directed telehealth parent-mediated intervention for children with autism spectrum disorder: Examination of the potential reach and utilization in community settings. *Journal of Medical Internet Research, 19*(7), e248. htpps://doi:10.2196/jmir.7484

Jacobs, K., Cason, J., & McCullough, A. (2015). The process for the formulation of the international telehealth position statement for occupational therapy. *International Journal of Telerehabilitation, 7*(1), 21-32.

Janes, W. E., Persch, A. S., Schwartz, J., & Cason, J. (2016). Current trends in assessment and technology. *SIS Quarterly Practice Connections, 1*(3), 12-14.

Kim, Y.-S., Al Otaiba, S., Sidler, J. F., Greulich, L., & Puranik, C. (2014). Evaluating the dimensionality of first grade written composition. *Journal of Speech, Language, and Hearing Research, 57*, 199-211.

Kleynen, M., Braun, S. M., Rasquin, S. M., Bleijlevens, M. H., Lexis, M. A., Halfens, J., & Beurskens, A. J. (2015). Multidisciplinary views on applying explicit and implicit motor learning in practice: An international survey. *PloS One, 10*(8), e0135522. https://doi: 10.1371/journal.pone.0135522

Knowles, C., Massar, M., Raulston, T. J., & Machalicek, W. (2017). Telehealth consultation in a self-contained classroom for behavior: A study. *Preventing School Failure: Alternative Education for Children and Youth, 61,* 28-38. doi:https://dx.doi.org/10.1080/1045988X.2016.1167012

Kuravackel, G. M., Ruble, L. A., Reese, R. J., Ables, A. P., Rodgers, A. D., & Toland, M. D. (2018). COMPASS for hope: Evaluating the effectiveness of a parent training and support program for children with ASD. *Journal of Autism and Developmental Disorders, 48*(2), 404-416. htpps://doi:10.1007/s10803-017-3333-8

Lautamo, T., Laakso, M., Aro, T., Ahonen, T., & Törmäkangas, K. (2011). Validity of the play assessment for group settings: An evaluation of differential item functioning between children with specific language impairment and typically developing peers. *Australian Occupational Therapy Journal, 58* (4), 222-230.

Lennon, C., & Burdick, H. (2004). *The lexile framework as an approach for reading measurement and success.* MetaMetrics. https://lexile.com/m/resources/materials/Lennon__Burdick_2004.pdf

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328. httpss://www.rasch.org/rmt/rmt74m.htm

Linacre, J. M. (2002). *Optimizing rating scale category effectiveness.* https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.2811&rep=rep1&typ

Linacre, J. M. (2010). When to stop removing items and persons in Rasch misfit analysis? *Rasch Measurement Transactions, 23,* 1241. https://www.rasch.org/rmt/rmt234g.htm

Linacre, J. M. (2020a). Winsteps® (Version 4.7.0) [Computer Software]. Winsteps.com. https://www.winsteps.com/

Linacre, J. M. (2020b). *Winsteps® Rasch measurement computer program manual 4.7.0 user's guide.* Winsteps.

Lincoln, Y., & Guba, E. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation, 20,* 15-25.

Little, L., Dunn, W., Pope, E., & Wallisch, A. (2016). Feasibility of a telehealth coaching intervention for families of children with autism. *American Journal of Occupational Therapy, 70,* 1-1. htpps://doi:10.5014/ajot.2016.70S1-PO1059

Lynn, M. R. (1998). Determination and quantification of content validity. *Nursing Research, 35,* 382-385.

Magill, R. A., & Anderson, D. I. (2014). *Motor learning and control: Concepts and applications* (10th ed.). McGraw Hill.

Magno, C. (2009). Demonstrating the difference between Classical Test Theory and Item Response Theory using derived test data. Online Submission. https://search-ebscohost

Masters, G. N. (1982). Constructing an item bank using partial credit scoring. *Journal of Educational Measurement, 21,* 19-32.

McAllister, S. (2008). Introduction to the use of Rasch analysis to assess patient performance. *International Journal of Therapy and Rehabilitation, 15*(11), 482-490.

McHale, K., & Cermak, S. A. (1992). Fine motor activities in elementary school: Preliminary findings and provisional implications for children with fine motor problems. *American Journal of Occupational Therapy, 46,* 898-903.

Milone, M. (2007). *Test of Handwriting Skills-Revised.* Academic Therapy Publications.

Mokkink, L. B., Prinsen, C. A., Bouter, L. M., Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based standards for the Selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy, 20*(2), 105-113. https://doi.org/10.1590/bjpt-rbf.2014.0143

Muratori, L. M., Lamberg, E. M., Quinn, L., & Duff, S. (2013). Applying principles of motor learning and control to upper extremity rehabilitation. *Journal of Hand Therapy, 26*(2), 94-103. https://doi.org/10.1016/j.jht.2012.12.007

Musselman, K. E., Roemmich, R. T., Garrett, B., & Bastian, A. J. (2016). Motor learning in childhood reveals distinct mechanisms for memory retention and re-learning. *Learning Memory, 23,* 229-237. htpps://doi: 10.1101/lm.041004.115

Nicola, K., Waugh, K., Charles, E., & Russell, T. (2018). The feasibility and concurrent validity of performing the Movement Assessment Battery for Children-2nd Edition via telerehabilitation technology. *Research in Developmental Disabilities, 77,* 40-48. htpps://doi.org/10.1016/j.ridd.2018.04.001

Norbakht, Z., Rassafiani, M., Hosseini, S. A., & Ahmadi, M. (2017). Telehealth in occupational therapy: A scoping review. *International Journal of Therapy and Rehabilitation, 24*(12), 534-538.

Olsen, J., & Knapton, E. (2006). *The Print Tool evaluation and remediation package*. Handwriting Without Tears.

Olson, C., McSwain, S. D., Curfman, A. L., & Chuo, J. (2018). The current pediatric telehealth landscape. *Pediatrics, 141*(3). htpps://doi:10.1542/peds.2017-2334

Palmis, S., Danna, J., Velay, J. L., & Longcamp, M. (2017). Motor control of handwriting in the developing brain: A review. *Cognitive Neuropsychology, 34*, 187-204.

Phelps, J., & Stempel, L. (1987). *Children's Handwriting Evaluation Scale for manuscript writing*. Texas Scottish Rite Hospital for Crippled Children.

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health, 30*, 459-467. https://dx.doi.org/10.1002/nur.20199

Reifenberg, G., Gabrosek, G., Tanner, K., Harpster, K., Proffitt, R., & Persch, A. (2017). Feasibility of pediatric game-based neurorehabilitation using telehealth technologies: A case report. *The American Journal of Occupational Therapy, 71*(3), 1-8. doi:https://dx.doi.10.5014/ajot.2017.024976

Reisman, J. (1999). *Minnesota Handwriting Test.* Psychological Corporation.

Robinson, R., & Penman, M. (2011). Viewing development through an occupational lens: Learning to handwrite. *New Zealand Journal of Occupational Therapy, 58*(2), 23-29.

Rosenbaum, D. A. (2009). *Human motor control* (2nd ed.). Elsevier. https://www.sciencedirect.com/science/book/9780123742261

Roston, K. L., Hinojosa, J., & Kaplan, H. (2008). Using the Minnesota Handwriting Assessment and handwriting checklist in screening first and second graders' handwriting legibility. *Journal of Occupational Therapy, Schools and Early Intervention, 1*(2), 100-115. https://doi.org/10.1080/19411240802312947

Ruble, L. A., McGrew, J. H., Toland, M. D., Dalrymple, N. J., & Jung, L. A. (2013). A randomized controlled trial of COMPASS web-based and face-to-face teacher coaching in autism. *Journal of Consulting and Clinical Psychology, 81*, 566.

Sangster-Jokić, C., & Whitebread, D. (2011). The role of self-regulatory and metacognitive competence in the motor performance difficulties of children with developmental coordination disorder: A theoretical and empirical review. *Educational Psychology Review, 23*(1), 75-98. https://www.jstor.org/stable/23883400

Schein, R.M., Schmeler, M.R., Holm, M.B., Pramuka, M., Saptono, A.& Brienza, D.M.(2011).Telerehabilitation assessment using the Functioning Everyday with a Wheelchair-Capacity instrument. *Journal of Rehabilitation Research and Development, 48*(2),115-24. doi:10.1682/jrrd.2010.03.0039.

Schwartz, Y., Averbuch, S., Katz, N., & Sagiv, A. (2016). Validity of the Functional Loewenstein Occupational Therapy Cognitive Assessment (FLOTCA). *The American Journal of Occupational Therapy, 70*(1), 1-7.

Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*, 205-231.

Smith, E. V., & Smith, R. M. (2004). *Introduction to Rasch measurement: Theory, models and applications*. JAM Press.

Smith, R. M., & Suh, K. K. (2003). Rasch fit statistics as a test of the invariance of item parameter estimates. *Journal of Applied Measurement, 4*(2), 153-163.

Staats, C., Oakley, G., & Marais, I. (2019). A legibility scale for early primary handwriting: Authentic task and cognitive load influences. *Issues in Educational Research, 29*(2), 537-556.

Stievano, P., Michetti, S., McClintock, S. M., Levi, G., & Scalisi, T. (2016). Handwriting fluency and visuospatial generativity at primary school. *Reading and Writing, 29*, 1497-1510.

Tomlinson, S. R. L., Gore, N., & McGill, P. (2018). Training individuals to implement applied behavior analytic procedures via telehealth: A systematic review of the literature. *Journal of Behavioral Education, 27,* 172. https://doi-org.tc.idm.oclc.org/10.1007/s10864-018-9292-0

Tseng, M. H., & Chow, S. M. K. (2000). Perceptual-motor function of school-age children with slow handwriting speed. *American Journal of Occupational Therapy, 54*(1), 83-88.

U.S. Department of Health and Human Services Office for Civil Rights. (2020, March 30). *Notification of enforcement discretion for telehealth remote communications during the COVID-19 nationwide public health emergency*. https://www.hhs.gov/hipaa/for-professionals/telehealth- index.html

van Hartingsveldt, M. J., de Groot, I. J. M., Aarts, P. B. M., & Nijhuis-van der Sanden, M. W. G. (2011). Standardized tests of handwriting readiness: A systematic review of the literature. *Developmental Medicine and Child Neurology, 53*(6), 506-515.

Veras, M., Kairy, D., Rogante, M., & Giacomozzi, C. (2015). Outcome measures in tele-rehabilitation and virtual reality for stroke survivors: Protocol for a scoping review. *Global Journal of Health Science, 8*(1), 79-82. https://doi.org/10.5539/gjhs.v8n1p79

Vismara, L. A., McCormick, C. E. B., Wagner, A. L., Monlux, K., Nadhan, A., & Young, G. S. (2018). Telehealth parent training in the Early Start Denver Model: Results from a randomized controlled study. *Focus on Autism and Other Developmental Disabilities, 33*(2), 67-79.

Wainer, A. L., & Ingersoll, B. R. (2015). Increasing access to an ASD imitation intervention via a telehealth parent training program. *Journal of Autism and Developmental Disorders, 45*(12), 3877-3890. https://doi:10.1007/s10803-014-2186-7

Wright, B. D. (1996a). Reliability and separation. *Rasch Measurement Transactions, 9*(4), 472. https://www.rasch.org/rmt/rmt94n.htm

Wright, B. D. (1996b). Local dependency, correlations, and principal components. *Rasch Measurement Transactions, 10* (3), 509. https://www.rasch.org/rmt/rmt191h.htm

Wühr, P., & Ansorge, U. (2019). Do left-handers outperform right-handers in paper-and-pencil tests of attention? *Psychological Research,* 1-11. doi:http://dx.doi.org.ezp.scranton.edu/10.1007/s00426-019-01224-z

Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research, 25*(5), 508-518. https://doi.org/10.1177/0193945903252998

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A. R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences, 4*(2), 165-178.

Ziegmann, M., Cole, E., Lichtenberg, A., & Brooks, R. (2001, November 25). Telerehabilitation: New tools for providing in-home brain injury rehabilitation. *OT Practice.*

Zwicker, J. G., & Harris, S. R. (2009). A reflection on motor learning theory in pediatric occupational therapy practice. *The Canadian Journal of Occupational Therapy, 76*(1), 29-37.

Appendix A

TeleWrite Assessment

TeleWrite© Handwriting Evaluation

(Web-based Observations)

Child's Name: _____ Sex: _____

Assessment Date: _____

Date of Birth: _____

Evaluator: _____

Setting: _____

Grade Level: _____

A. Functional Observations

Directions: Virtually observe the child at home and/or in the classroom. Please indicate your response by checking the selected item. **Complete this section over telehealth**

Writing Utensil:

_____ Standard pencil (#2) _____ Primary pencil (wide pencil)

Adaptations:

_____

Use of a Visual Aid: (i.e. glasses or other aids) Y N

If yes, what visual aid was used? _____

Handedness:

_____ Right _____ Left _____ Mixed

Stabilization with non-writing hand:

_____ Yes _____ No _____ Sometimes

Pressure on Paper:

a. Wavy or light lines Y N
b. Excessive pressure (Piercing paper) Y N

B. In- Hand Manipulation Assessment

Directions: Virtually observe the child at home and/or in the classroom. Please indicate your response by using the score guide below. **Complete this section over telehealth**

Coin:

A. Finger to palm translation with stabilization using a coin

☐ Skill is Present ☐ Skill is Emerging, partial skill is present ☐Skill is Absent

B. Palm to finger translation with stabilization using a coin

☐ Skill is Present ☐ Skill is Emerging, partial skill is present ☐Skill is Absent

Pencil:

A. Simple rotation of a pencil – tip placed ulnar (pinky) side (picks up pencil from surface using thumb, index, and third digit.)

☐ Skill is Present ☐ Skill is Emerging, partial skill is present ☐Skill is Absent

B. Complex rotation of pencil after tip is oriented to radial (thumb) side (being able to turn pencil on its end to erase)

☐ Skill is Present ☐ Skill is Emerging, partial skill is present ☐Skill is Absent

C. Shift on pencil

☐ Skill is Present ☐ Skill is Emerging, partial skill is present ☐Skill is Absent

Pencil Grasp: Please indicate which grasp is observed.

_____Dynamic tripod _____ Static tripod _____ Quadrupod

_____Stenographer _____ Thumb Wrap _____ Thumb tuck

_____ Lateral Tripod _____ Other

Comments on atypical grasp and webspace:

_____

Adapted from: Case-Smith & Exner, C (2015) Hand Function Evaluation and Intervention. In J. Case-Smith & J.C. O'Brien (Eds.), *Occupational therapy for children* (7th ed) (pp. 225-226). St. Louis: Mosby.

C. Posture and Behavioral Observation

Directions: Virtually observe the child at home and/or in the classroom *while* the child writes upper and lowercase letters. Please make observations below. (*Note that this section is not timed.)* **Complete this section over telehealth**

Tabletop Sitting Posture:

_____

_____

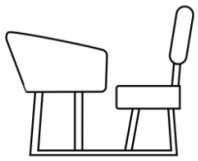
(i.e., slouching, leaning on UE, sitting up straight, etc.)

Fidgeting and Behavior Patterns:

_____

_____

(Describe fidgeting behaviors and attitude i.e., shaking, squirming, complaining, etc.)



D. Handwriting Analysis of Completed Forms

Directions: Score *after* receiving mail packet from parent/teacher. Please indicate your response in the section below and calculate the accuracy.
Complete this section after receiving completed packet

Task 1: Rote Letter Formation: Upper Case Letters

Score Guide**:**
Score 0: Letter is not recognizable out of context, or is reversed
Score 1: Letter is somewhat recognizable in context
Score 2: Letter clearly resembles the model

A_____	F_____	K_____	P_____	U_____	Z_____
B_____	G_____	L_____	Q_____	V_____	
C_____	H_____	M_____	R_____	W_____	
D_____	I_____	N_____	S_____	X_____	
E_____	J_____	O_____	T_____	Y_____	

TOTAL _____/52 (Maximum Points)

Rote Letter Uppercase Accuracy Formula: (Total ÷ 52) x100 = _____%
(round to the nearest hundredth)

Task 2: Rote Letter Formation: Lower Case Letters

Score Guide**:**
Score 0: Letter is not recognizable out of context, or is reversed
Score 1: Letter is somewhat recognizable in context.
Score 2: Letter clearly resembles the model.

a_____	f_____	k_____	p_____	u_____	
b_____	g_____	l_____	q_____	v_____	
c_____	h_____	m_____	r_____	w____	
d_____	i_____	n_____	s_____	x_____	
e_____	j_____	o_____	t_____	y_____	z_____

TOTAL _____/52 (Maximum Points)

Rote Letter Lower Case Accuracy Formula: (Total ÷ 52) x100 = _____%
(round to the nearest hundredth)

Task 3: Rote Numeral Formation

Score Guide**:**

Score 0: Number is not recognizable out of context or is reversed.

Score 1: Number is somewhat recognizable in context.

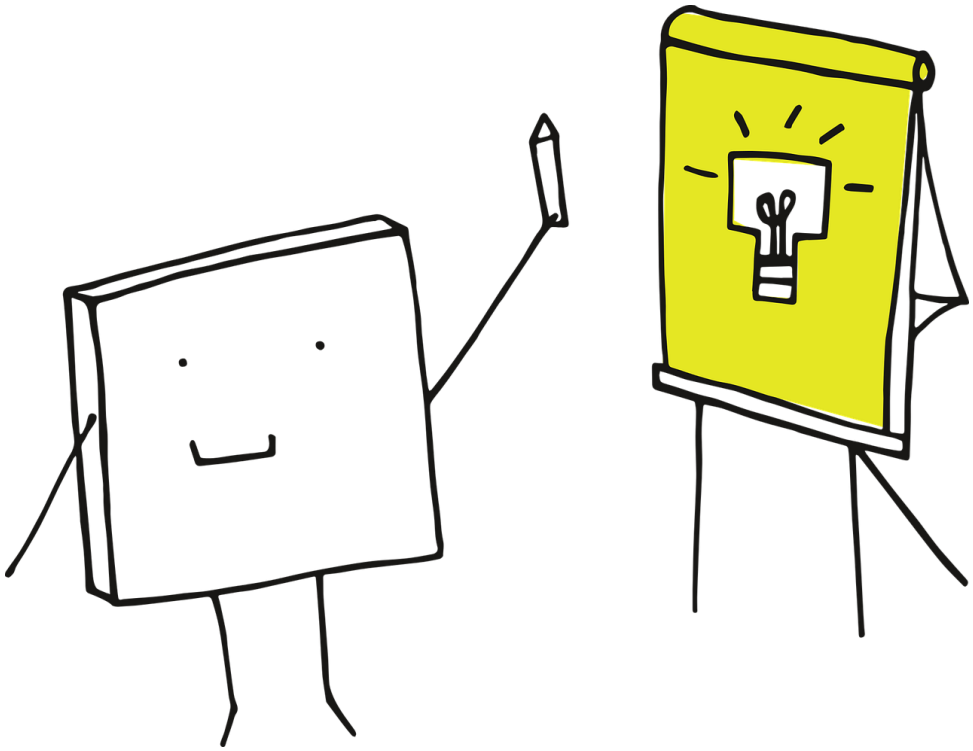
Score 2: Number clearly resembles the model.

1_____ 3_____ 5_____ 7_____ 9_____

2_____ 4_____ 6_____ 8_____ 0____

TOTAL_____/20 (Maximum Points)

Rote Numeral Accuracy Formula: (Total ÷ 20) x100 = _____%
(round to the nearest hundredth)



Task 4: Near Point Writing Prompts

Time to Complete:

Directions: Indicate the time taken to complete this section in the space provided. ________ seconds **Complete this section after receiving completed packet**

Accuracy:

Directions: Fill in the boxes with the number of letters in each word that are recognizable within context. Then, fill in the blank for the total number of correct letters.

/3	/4	/6	/7	/4	/7

The five boxing wizards jump quickly.

Correct Letters: ___/ 31

Near Point Accuracy Formula: (Total ÷ 31) x100 = _____% (round to the nearest hundredth)

Task 5: Far Point Writing Prompts

Time to Complete:

Directions: Indicate the time taken to complete this section in the space provided.
________ seconds **Complete this section after receiving completed packet**

Accuracy:

Directions: Fill in the boxes with the number of letters in each word that are recognizable within context. Then, fill in the blank for the total number of correct letters.

/3	/4	/3	/6	/6	/4	/1	/3	/4

Max, Jack, and Harvey helped with a big quiz.

Correct Letters: ___/ 34

Far Point Accuracy Formula: (Total ÷ 34) x100 = _____% (round to the nearest hundredth)

Task 6: Observation of Spatial Relationships in Sentence Composition

Directions: Review the student's near point and far point written samples. Further testing is recommended for scores that are 0 or 1. **Complete this section after receiving completed packet**

Score 0: Skill is observed in less than approximately 25% of the sample.
Score 1: Skill is observed in approximately 25% - 75% of the sample.
Score 2: Skill is observed in more than approximately 75% of the sample.

1. Letters are properly formed (closed, no open loops or gaps) _____

2. Letters are aligned proportionally to the line _____

3. Uniform spacing between letters (between ⅛ inch and ¼ inch space) ___

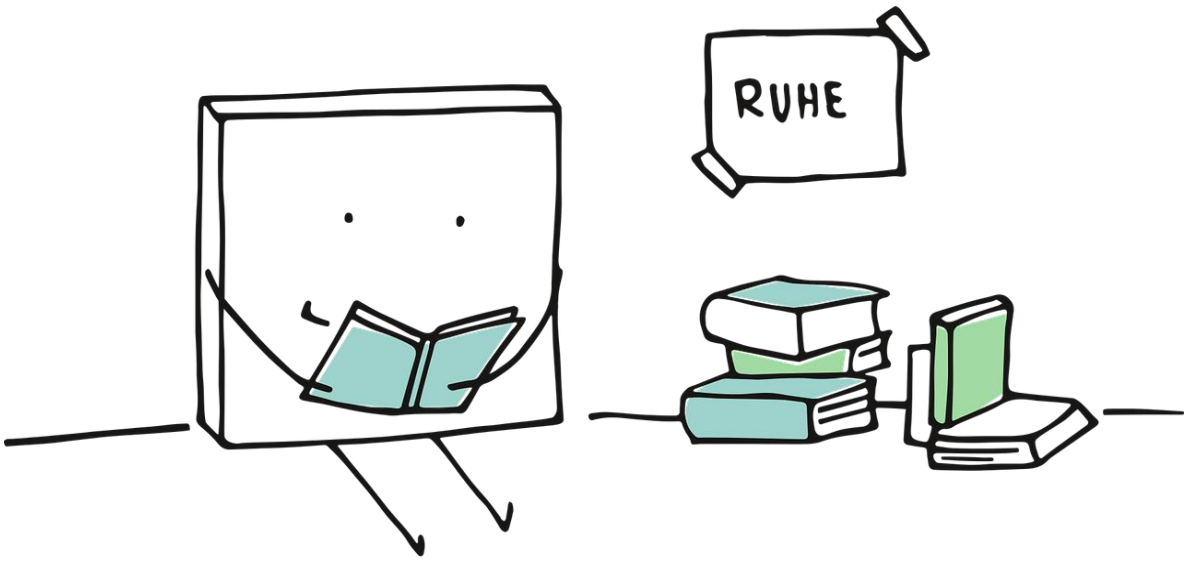
4. Uniform spacing between words (between ¼ inch and ½ inch space) _____

5. Letters are the same size ___

6. Descending letters fall appropriately below line ___

Observations: _____

TOTAL _____/12 points (Maximum Points)



E. TeleWrite Score Form

Handwriting Accuracy Scores

Directions: Move accuracy % calculations here. Round to the nearest hundredth.
Complete this section after receiving completed packet

Upper case letter accuracy (p. 4)	**%**
Lower case letter accuracy (p. 4)	**%**
Numeral accuracy (p. 5)	**%**
Near point sentence accuracy (p. 6)	**%**
Far point sentence accuracy (p. 7)	**%**

Rating Scale for Near Point Accuracy Directions: Circle score that child achieved 1-10.99% Unsatisfactory 11-20.99% - Very lo 21-30.99% - Low 31-40.99% - Below Average 41-50.99% - Low Average 51-60.99% - Average 61-70.99% - Above Average 71-80.99% - High 81-90.99% - Very High 91-100% - Exceptional	**Rating Scale for Far Point Accuracy** Directions: Circle score that child achieved 1-10.99% Unsatisfactory 11-20.99% - Very low 21-30.99% - Low 31-40.99% - Below Average 41-50.99% - Low Average 51-60.99% - Average 61-70.99% - Above Average 71-80.99% - High 81-90.99% - Very High 91-100% - Exceptional

Timing

Near Point: _____ seconds ÷ 60 = _____ minutes

Far Point: _____ seconds ÷ 60 = _____ minutes

Writing Rate - (Letters per minute)

Near Point Writing Rate: 31 ÷ _____ minutes = _____ letters per minute

Far Point Writing Rate: 34 ÷ _____ minutes = _____ letters per minute

Writing Fluency - (Legible letters per minute)

Near Point Fluency: _____ number of correct letters (p. 6) ÷ _____ minutes = _____ legible letters per minute

Far Point Fluency: _____ number of correct letters (p. 7) ÷ _____ minutes = _____ legible letters per minute

Adapted from:

Mean Handwriting Speeds: Letters per Minute		
	Girls	Boys
Grade 1	20.51	17.44
Grade 2	36.77	31.55
Grade 3	49.80	44.80

Graham, S., Berninger, V. W.& Weintraub, N. (1998). Development of handwriting speed and legibility in grades 1-9. *Journal of Educational Research, 92*, 42-52.

G. Additional Notes/Observations

Appendix B

TeleWrite Script

Occupational Therapist Script (To Read to Parent)

"Thank you for participating in our trial of the TeleWrite, a handwriting assessment delivered online. Before we begin, please look through and familiarize yourself with the packet."

Introduction

Please follow all directions carefully and use the sentences or phrases as a guide. All commands written in italics are for you, the occupational therapists, to do, and not to be said aloud.

Introduce yourself to the parent and the child and confirm the child's grade level (1, 2, or 3). This section does not require you to relay any specific information.

Setting Up the Environment

(Instructions for caregiver/teacher)

"Before we begin, please set up the environment to avoid distractions during the test. Please look around to note what may need to be changed in the environment. (i.e. have pets in another room, have other children in another room). Please make sure that the computer is facing the child."

Assist caregiver/teacher to make sure the child is able to see the therapist on the screen.

- Before we begin, please make sure that the child has a pencil that they usually use to complete writing assignments.
- "Open up your packet to the section marked by the child's grade level. Do not turn the page until you are instructed to do so."
- "Take off the last page in the packet." Can you show me the page?

  Page number will vary depending on the child's grade level. Indicate page number to caregiver or teacher.

- "Please set up the screen so I can see the child's hands and face. This will be approximately 4' from the child's desk. This is where you will stand and hold the paper for the child to copy."
- "Do you have all the materials?" (coin, pencil, booklet)
- "Do you have any questions before we begin?"

PART A. Functional Observations & PART B. In-hand Manipulation Assessment

- "Remove the coin from the tape and place on desk in front of the child."
- *Direct the child to bring coin to his/her palm.* "Pick up the coin using the fingertips of your writing hand without touching your body for help."
- *Direct the child to bring coin from palm to his fingertips- 1st, 2nd digits and thumb.* "Now, move the coin in your hand to your fingertips."
- "Show me the coin in your palm" *May demonstrate if needed.*
- "Please place a pencil on the table with the tip toward the child's pinky finger."

Direct the child to pick up the pencil in simple rotation- may demonstrate if necessary.

"Pick up the pencil as if you are going to write."

- "Please place a pencil on the table with the tip toward the thumb side."

Direct the child to pick up the pencil in complex rotation- may demonstrate if necessary.

- "Please have the child take the pencil in his writing hand with the tip facing down. *Direct the child to "shift" fingers on pencil- may demonstrate if necessary.*
- "Walk your fingers up and down the shaft of the pencil without dropping it."

First Grade

As you are administering the assessment, please note the child's posture and behavior (PART C on scoring sheet).

PART D of the assessment includes the handwriting analysis of the child's completed forms. This includes rote letter formation of uppercase letters (TASK 1), lower case letters (TASK 2), and numbers (TASK 3).

TASK 4 on the scoring sheet includes Near Point Writing Prompts:

Near Point Alphabet

"Please turn the page to 2 in your packet. When I say go, please print all of the capital letters on the lines starting with the letter S. Use the space provided on lines to write your answers. Please do this as quickly and as neatly as you can. There is no erasing on this test. You may begin when I say go. Go."

On the next page there are lowercase letters. Please copy the letters on the lines below as quickly and as neatly as you can when I say go. Go."

"On the next page, please copy the numbers in order, as quickly and neatly as possible. You may begin when I say go. Go."

*When the child is finished, "*Can you quickly show me your work?

Near Point Sentence Writing Prompt

"Please turn to page 5 in your packet. Please read aloud the sentence to me. Copy the sentence in your packet in the space provided below. Remember to do this as quickly and neatly as you can. When you are finished say "done" so I will know you are finished. You may begin when I say go. Go."

Start the timer as soon as the child picks up the pencil. Stop the timer as soon as the child finishes the last letter. Record the correct time in seconds on the line labeled 'Time to Complete' located under Task 4 of the scoring sheet.

TASK 5 on the scoring sheet includes the Far Point Writing Prompt

Far Point Sentence Writing Prompt

"Please turn to page 6 in your packet. Please have the child read the prompt aloud. Please stand behind the computer/ iPad screen (approximately 2 feet away) from the child and hold the paper at the child's eye level and away from the ground. On your lined sheet of paper please copy what you see on the sheet as quickly and as neatly as you can when I say go. When you are finished say "done" so I will know you are finished. Go."

Start the timer as soon as the child picks up the pencil. Stop the timer as soon as the child finishes the last letter. Record the correct time in seconds on the line labeled on the line labeled 'Time to Complete', located under Task 5 on the scoring sheet.

Conclusion

This is the end of the assessment.

"You have now completed this handwriting assessment. Thank you for participating in this research study. Please don't forget to complete the parent survey and have the child complete the child survey. Please send the assessment back in the envelope provided within 3-5 days. Once we receive your packet in the mail, we will send a link through your email for additional feedback to help us improve the test. Do you have any questions? Thank you!"

Task 6 on the scoring sheet is the observation of spatial relationships in the sentence composition. PART E of the scoring sheet is the computation of scores.

Appendix C

Clinical Utility Survey

Q1 Thank you for participating in this survey. We are interested in assessing clinical utility of a new occupational therapy assessment tool, the TeleWrite. We are looking for OTs who: Have 3 years of experience administering pediatric handwriting assessments, currently have a valid OT license and have access to the internet.

You will be presented with 14 questions relevant to handwriting and the TeleWrite assessment. Please answer to the best of your ability. Your responses are completely confidential. This survey should take no more than 15 minutes to complete. Your participation in this research is voluntary. You have the right to withdraw at any point during the survey, for any reason and without any prejudice.

If you would like to contact the principal investigator of this study, Julia M. Guzman OTD, OTR/L, to discuss the research, please email Telewrite@gmail.com.

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you can terminate your participation in the study at any time and for any reason. This study was approved by IRB on 11/15/2019.

Q3 Attached is the first-grade version of the TeleWrite assessment. Second and third grade versions have also been developed with age-appropriate writing prompts. Please

review the first-grade assessment and respond to the survey.

1st grade TeleWrite 2020 2 6 1 copyright

Q4 In which country do you currently reside?

▼ Afghanistan (1) ... Zimbabwe (1357)

Q5 In which state do you currently reside?

▼ Alabama (1) ... I do not reside in the United States (53)

Q6 How many years have you been a practicing occupational therapist?
3-5 years (3)

6-10 years (4)

11-15 years (5)

More than 16 years (6)

Q7 In what setting do you primarily practice.

Academia (1)

Hospital Non-Mental Health (2)

Community (3)

Early Intervention (4)

Outpatient (5)

Home-health (6)

LTC/SNF (7)

Mental Health (8)

Schools (10)

Other please explain: (9) _____

Q8 Please select any other setting(s) where you practice.

Academia (1)

Hospital Non-Mental Health (2)

Community (3)

Early Intervention (4)

Outpatient (5)

Home-health (6)

LTC/SNF (7)

Mental Health (8)

Schools (10)

Other please explain: (9) _____

N/A I only practice in the area mentioned in the prior question (11)

Q9 Do you have any experience with telehealth?

No telehealth experience at all (1)

Slightly experienced with telehealth (2)

Moderately experienced with telehealth (3)

Quite experienced with telehealth (4)

Extremely experienced with telehealth (5)

Q10 How comfortable are you with using technology (computers, iPad, etc.)?

Not comfortable at all (1)

Slightly comfortable (2)

Moderately comfortable (3)

Quite comfortable (4)

Extremely comfortable (5)

Q11 How easy are the instructions regarding the administration of the TeleWrite Assessment?

Not at all easy (1)

Slightly easy (2)

Moderately easy (3)

Quite easy (4)

Extremely easy (5)

Q12 How can the instructions on the TeleWrite assessment be enhanced?

_____

Q13 How accurately does the TeleWrite assessment measure handwriting performance?

Not at all accurate (1)

Slightly accurate (2)

Moderately accurate (3)

Quite accurate (4)

Extremely accurate (5)

Q14 How can the accuracy of the TeleWrite assessment be enhanced?

_____

Q15 Based on your review of the TeleWrite assessment, how likely are you to use it in your practice?

Not at all likely (1)

Slightly likely (2)

Moderately likely (3)

Quite likely (4)

Extremely likely (5)

N/A This instrument is not applicable to my area of practice (6)

Q16 Please provide any additional feedback about the TeleWrite assessment.

_____

Q17 If you would like any updates about the development of this assessment please provide your email below.

_____

Appendix D

Content Validity Survey

Q2 Dear Content Expert Reviewer,

We are inviting you to participate in a study for a new telehealth-based handwriting tool for children in grades 1-3. The TeleWrite is a performance-based tool of handwriting delivered via Telehealth. If you choose to participate in this study, you will find an electronic copy of the TeleWrite on the next page via a link. We ask that you spend 30-45 minutes to thoroughly review each of the items and scales of the tool. You will also find a review survey with a 4-point scale asking you to review how relevant each of the items are. You will score whether each item and its scales as either: (1) not relevant, (2) minimally relevant, (3) moderately relevant, and (4) highly relevant. We will also provide a space where you can provide any comments about the item and scale in the TeleWrite. This process may take you another 30-45 minutes to complete. Total time to review the inventory will be 90 minutes.

A total of 10 experts will be recruited to participate in this study and must meet the following criteria: have ten or more years of clinical experience addressing handwriting or assessment with school-aged children, published a peer reviewed article in pediatrics, or have done research in instrument development.

If you have questions or concerns, please email Julia M. Guzman, OTD, OTR/L at jg3740@tc.columbia.edu or Dr. Lenin Grajo at lg2890@cumc.columbia.edu

By clicking the button below, you acknowledge that your participation in the study is voluntary, you are 18 years of age, and that you are aware that you can terminate your participation in the study at any time and for any reason. IRB Protocol Number 20-223

Q3 Do you consent to participate in this study?

o I consent (1)

o I do not consent (2)

Q4 How many years have you practiced as a pediatric occupational therapist? Q5 Please provide your highest academic degree.

o MS, MA, MOT (1)

o OTD (2)

o PhD, Ed.D. or another research doctorate (3)

Q5 Please provide approximate number of peer reviewed, chapter reviews and non-peer reviewed publications related to handwriting, telehealth, or pediatrics.

o 1-5 (1)

o 6-10 (2)

o 11-15 (3)

o 16+ (4)

Q6 Please provide approximate number of workshops, posters, short courses or in-service presentations related to handwriting, telehealth or pediatrics.

- 1-5 (1)
- 6-10 (2)
- 11-15 (3)
- 16+ (4)

Q7 In order to answer the questions in the next page, please click the link below to access the content reviewer version of the TeleWrite.

Q8 Please review "Near and Far Point Copying" section on page 2 and rate how relevant this item is in assessing handwriting. We encourage you to add comments so we can improve this section.

Not relevant (1)

Minimally relevant (2)

Moderately relevant (3)

Highly relevant (4)

Q9 Please give any comments on how we can improve "Near and far Point Copying."

Q10 Please review the "Rote Letter Formation" and "Rote Numeral Formation" on page 3 and rate how relevant these items are in assessing handwriting. We encourage you to add comments so we can improve this section.

Not relevant (1)

Minimally relevant (2)

Moderately relevant (3)

Highly relevant (4)

Q11 Please give any comments on how we can improve "Rote Letter Formation" and "Rote Numeral Formation" on page 3.

Q12 Please review the "Near Point Writing Prompts" on page 4 and rate how relevant these items are in assessing handwriting. Please note, TeleWrite prompts are made for each of the individual grades 1-3. For your review, we listed all grade prompts here. We encourage you to add comments so we can improve this section.

Q13 Please give any comments on how we can improve "Near Writing Prompts" on page 4.

Q14 Please review "Far Point Writing Prompts" on page 5 and rate how relevant these items are in assessing handwriting. Please note, TeleWrite prompts are made for each of

the individual grades 1-3. For your review, we listed all grade prompts here. We encourage you to add comments so we can improve this section.

Q15 Please give any comments on how we can improve "Far Point Writing Prompts."

Q16 SPATIAL RELATIONSHIPS in the TeleWrite is defined as visual appearance of writing sample. Closed, aligned, and spaced uniformly.

Please review "Spatial Relationships in Sentence Composition" on page 6 and rate how relevant these items are in assessing handwriting. We encourage you to add comments so we can improve this section.

o Not relevant (1)

o Minimally relevant (2)

o Moderately relevant (3)

Highly relevant (4)

Q17 Please give any comments on how we can improve "Spatial Relationships" on page 6.

Q18 TIME: In this section you will convert the speed of writing written in seconds to minutes. Please review "Timing" on page 7 and rate how relevant this formula is in assessing handwriting. We encourage you to add comments so we can improve this section.

- Not relevant (1)
- Minimally relevant (2)
- Moderately relevant (3)
- Highly relevant (4)

Q19 Please give any comments on how we can improve "Timing" formula on page 7.

Q20 WRITING RATE in the TeleWrite is calculated based on the number of letters written by the student in minutes. Please review "Writing Rate" on page 8 and rate how relevant this formula is in assessing handwriting. We encourage you to add comments so

we can improve this section. *Please note that what you are reviewing are the number of letters for first grade only. The number of letters will differ from second and third grade*

- Not relevant (1)
- Minimally relevant (2)
- Moderately relevant (3)
- Highly relevant (4)

Q21 Please give any comments on how we can improve "Writing Rate" formula on page 8.

Q22 WRITING ACCURACY in the TeleWrite is calculated as the number of letters written CORRECTLY per minute.

Please review the "Handwriting Accuracy Scores" on page 9 and rate how relevant these formulas are in assessing handwriting. We encourage you to add comments so we can improve this section.

- Not relevant (1)
- Minimally relevant (2)
- Moderately relevant (3)
- Highly relevant (4)

Q23 Please give any comments on how we can improve "Writing Accuracy" formula on page 10.

Q24 WRITING FLUENCY in the TeleWrite is calculated based on how FAST (rate) a student can CORRECTLY (accuracy) copy the letters per minute. Please review "Writing Fluency Formula" on page 10 and rate how relevant these formulas are in assessing handwriting.

- Not relevant (1)
- Minimally relevant (2)
- Moderately relevant (3)
- Highly relevant (4)

Q25 Please give any comments on how we can improve "Writing Fluency" formula.

Q26 Please review each of the prompts on page 11 (near point) and page 12 (far point). Please provide any comments or feedback.

Appendix E

Data Collection Form

Therapist Time Log

Child's Initials: _____

Packet #: _____

Grade Level: _____

Time Required

Near Point Copy: _____ seconds

Notes: _____

Far Point Copy: _____ seconds

Notes: _____